A robust approach for estimating change-points in the mean of an $AR(p)$ process

Souhil Chakar[*†]
souhil.chakar@imag.fr

September 4, 2015

Abstract

We consider the problem of multiple change-point estimation in the mean of an AR(p) process. Taking into account the dependence structure does not allow us to use the inference approach of the independent case. Especially, the dynamic programming algorithm giving the optimal solution in the independent case cannot be used anymore. We propose a two-step method, based on the preliminary estimation of the autoregression parameters. It is based on robust statistics techniques, since our estimator has to be robust to the change-points if we do not want to estimate them before. Then, we propose to follow the classical inference approach, by plugging this estimator in the criterion used for change-point estimation, which is equivalent to decorrelate the series using the estimated autoregression parameters. We show that the asymptotic properties of these change-point location and mean estimators are the same as those of the classical estimators in the independent framework. The same plug-in approach is then used to approximate the modified BIC and choose the number of segments, and to derive a heuristic BIC criterion to select both the number of changes and the order of the autoregression. Finally, we show, in the simulation section, that for finite sample size taking into account the dependence structure improves the statistical performance of the change-point estimators and of the selection criterion.

1 Introduction

Change-point detection problems arise in many fields, such as genomics (Braun and Müller, 1998; Braun et al., 2000; Picard et al., 2005), medical imaging (Lavielle, 2005), earth sciences (Williams, 2003; Gazeaux et al., 2013) or climate (Mestre, 2000; Lu et al., 2010). In many of these problems, the observations cannot be assumed to be independent.

In the literature, in the frequentist framework, there is two ways to deal with the dependence structure of series affected by multiple change-points:

- Apply the methodology of the independent case, and prove that asymptotic results are not affected by dependence under some conditions (Lavielle and Moulines, 2000; Lavielle, 1999).
- Consider that all the parameters of the model can change at each change-point (Bardet et al., 2012). In fact, inference is performed by considering also possible changes in the spectrum of the process.

[*]INP Grenoble, Laboratoire Jean Kuntzmann, Grenoble cedex 09, France.
[†]INRA-AgroParisTech, UMR 518 MIA-Paris, Paris cedex 05, France.

The parameters of the dependence structure are here assumed to be global (i.e. not depending on the segments) nuisance parameters that have to be estimated to perform the change-points inference.

In this paper, we consider the segmentation of an AR(p) process with homogeneous autoregression coefficients:

$$y_i = \mu_k^\star + \eta_i \ , \ t_{n,k}^\star + 1 \le i \le t_{n,k+1}^\star \ , \ 0 \le k \le m^\star \ , \ 1 \le i \le n \ , \tag{1}$$

where $(\eta_i)_{i\in\mathbb{Z}}$ is a zero-mean stationary AR(p) process. That is, it is a stationary solution of

$$\forall i \in \mathbb{Z}, \eta_i - \sum_{r=1}^{p} \phi_r^\star \eta_{i-r} = \epsilon_i \ , \tag{2}$$

where the ϵ_i's are uncorrelated zero-mean rv's with variance σ^2 and $\Phi^\star = (\phi_r^\star)_{1\le r\le p} \in \mathbb{R}^p$ is such that a stationary solution to (2) exists. In (1), we use the following conventions: $t_{n,0}^\star = 0, t_{n,m^\star+1}^\star = n$.

Our aim is to propose a methodology for estimating both the change-point locations $\boldsymbol{t_n^\star} = (t_{n,k}^\star)_{1\le k\le m^\star}$ and the means $\boldsymbol{\mu}^\star = (\mu_k^\star)_{0\le k\le m^\star}$, accounting for the existence of $\Phi^\star$. Moreover, we propose a criterion to choose the number of change-points $m^\star$.

In the sequel, we shall assume that there exists $\boldsymbol{\tau}^\star = (\tau_k^\star)_{0\le k\le m+1}$ such that, for $0 \le k \le m+1$ $t_{n,k}^\star = \lfloor n\tau_k^\star \rfloor$, $\lfloor x \rfloor$ denoting the integer part of x. Consequently, $\tau_0^\star = 0$ and $\tau_{m^\star+1}^\star = 1$.

A first idea is to use the maximum-likelihood approach to estimate the parameters of the model. However, maximizing the likelihood function, especially in the change-point location parameters $\boldsymbol{\tau}$, leads to a complex discrete optimization problem in an algorithmic point of view.

When the observations are independent, the optimal segmentation (e.g. in the maximum-likelihood sense) can be recovered via the dynamic programming (DP) algorithm introduced by Auger and Lawrence (1989). The computational complexity of this algorithm is quadratic relatively to the length of the series. This algorithm and some of its improvements (such as these proposed by Rigaill, 2010; or Killick et al., 2012) are the only one that provide exactly the optimal change-point location estimators. However, the DP algorithm only applies when (i) the loss function (e.g. the negative log-likelihood) is additive with respect to the segments and when (ii) no parameter to be estimated is common to several segments.

In the autoregressive case, the likelihood function violates both requirements (i) and (ii). Indeed the log-likelihood is not additive with respect to the segments because of the dependence that exists between data from neighbor segments and the unknown coefficients $\Phi^\star$ needs to be estimated jointly over all segments. Even if $\Phi^\star$ was known, the DP principle would not apply to the log-likelihood of Model (1) as it will still not be additive. We introduce an alternative criterion, based on the quasi-likelihood described by Bardet et al. (2012). This criterion is equivalent to the classic least-squares applied to a decorrelated version of the series, computed with an estimated $\widehat{\Phi}$. To achieve this decorrelation, we shall provide an estimator of $\Phi^\star$.

We shall prove that, under mild asymptotic assumptions on the estimator of $\Phi^\star$, the resulting change-point estimators satisfy the same rate of convergence as those proposed by Lavielle and Moulines (2000); Bardet et al. (2012).

We show that such an estimator of $\Phi^\star$ exists and can be computed before segmenting the series. In order to estimate $\Phi^\star$, we first differentiate the series of observations (y_i) and work on $x_i = y_i - y_{i-1}$ which satisfies

$$x_i = \nu_i \text{ except for } i = t_{n,k}^\star + 1, \text{ where } x_{t_{n,k}^\star+1} = (\mu_k^\star - \mu_{k-1}^\star) + \nu_{t_{n,k}^\star+1}, \ 0 \le k \le m^\star \ , \tag{3}$$

where $\nu_i = \eta_i - \eta_{i-1}$ is an ARMA(p,1) defined from (η_i) by

$$\nu_i - \sum_{r=1}^{p} \phi_r^\star \nu_{i-r} = \epsilon_i - \epsilon_{i-1} \; . \tag{4}$$

To this aim, we borrow techniques from robust estimation (Ma and Genton, 2000). Briefly speaking, we consider the data observed at the change-point locations as outliers and propose an estimator of $\Phi^\star$ that is robust to the presence of such outliers. We shall prove that the estimator that we propose is consistent and asymptotically Gaussian. Moreover, we propose a model selection criterion on the number of changes, the order of the autoregression being fixed, inspired by the one proposed by Zhang and Siegmund (2007) and prove some asymptotic properties of this criterion. Finally, we discuss the problem of selecting jointly the number of changes and the order of the autoregression and propose a practical solution to this problem based on a Bayesian heuristic.

2 Robust estimation of the autoregression coefficients

The aim of this section is to provide an estimator of $\Phi^\star$ which can deal with the presence of change-points in the data. In the absence of change-points ($m^\star = 0$ in (1)), the estimation of $\Phi^\star$ is a well-known issue (see Brockwell and Davis, 1991, for a comprehensive introduction) and a consistent, asymptotically Gaussian estimator is given by the Yule-Walker method. We aim to adapt this method to our framework.

Since change-points can be seen as outliers in the AR(p) process, we shall propose a robust approach for estimating $\Phi^\star$. Our approach is based on the estimator of the autocorrelation function of a stationary time series proposed by Ma and Genton (2000) which is based on the robust scale estimator of Rousseeuw and Croux (1993). More precisely, let us define $\widetilde{\Phi}_n^{(p)}$ by

$$\widetilde{\Phi}_n^{(p)} = \widetilde{R}_{n,p}^{-1} \widetilde{\boldsymbol{\rho}}_{n,2:(p+1)}^T \; , \tag{5}$$

where for $i < j$ integers,

$$\widetilde{\boldsymbol{\rho}}_{n,i:j} \quad = \quad (\widetilde{\rho}_n(h))_{i \le h \le j} \; , \tag{6}$$

is an estimator of $\boldsymbol{\rho}_{i:j}$ defined by

$$\boldsymbol{\rho}_{i:j} \quad = \quad (\rho(h))_{i \le h \le j} \; , \tag{7}$$

where $\rho(h)$ denotes the autocorrelation of the process (x_i) defined in (3) at lag h, and $\cdot^T$ state for the transpose. In (5), $\widetilde{R}_{n,p}$ denotes the following matrix

$$\widetilde{R}_{n,p} \quad = \quad (\widetilde{\rho}_n\,(j-i-1))_{1 \le i,j \le p} \; , \tag{8}$$

which is an estimator of

$$R_p \quad = \quad (\rho\,(j-i-1))_{1 \le i,j \le p} \; . \tag{9}$$

Moreover, for all h in $\mathbb{Z}$,

$$\widetilde{\rho}_n(h) = \frac{Q_n^2\left(x_h^+\right) - Q_n^2\left(x_h^-\right)}{Q_n^2\left(x_h^+\right) + Q_n^2\left(x_h^-\right)} \; , \tag{10}$$

where $x_h^+ = (x_{i+h} + x_i)_{0 \le i \le n-h}$, $x_h^- = (x_{i+h} - x_i)_{0 \le i \le n-h}$ and Q_n is the scale estimator of Rousseeuw and Croux (1993) which is such that $Q_n\,(z)$ is proportional to the first quartile of $\{|z_i - z_j|; 0 \le i < j \le n\}$.

Proposition 2.1. *Let* $\widetilde{\boldsymbol{\rho}}_{n,i:j} = (\widetilde{\rho}_n(h))_{i\le h\le j}$ *and* $\boldsymbol{\rho}_{i:j} = (\rho(h))_{i\le h\le j}$ *be defined in* (6), (7) *and* (10). *Under the assumption that* ν_i *in* (3) *is Gaussian, we have*

$$n^{1/2}\left(\widetilde{\boldsymbol{\rho}}_{n,1:(p+1)} - \boldsymbol{\rho}_{1:(p+1)}\right) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0,V) \tag{11}$$

in distribution, where $\mathcal{N}(0,V)$ *is the* $(p+1)-$ *dimensional centered Gaussian distribution with covariance matrix* V *and*

$$n^{1/2}\left(\widetilde{\Phi}_n^{(p)} - \Phi^\star\right) \underset{n\to\infty}{\longrightarrow} \mathcal{N}\left(0, M^T\left(R_p^{-1}\right)^T V R_p^{-1} M\right) , \tag{12}$$

in distribution, where

$$M = \left(\phi^\star_{i-j+1}\mathbf{1}_{i\ge j} + \phi^\star_{i+j+1}\mathbf{1}_{i+j\le p-1} - \mathbf{1}_{j=i+1}\right)_{1\le i\le p, 1\le j\le p+1} . \tag{13}$$

The proof of Proposition 2.1 is given in Section 6.

Remark 2.1. Proposition 2.1 argues for the possibility to estimate robustly $\Phi^\star$. Equations (12) and (13) show that the variance of $\widehat{\Phi}$ depends on $\Phi^\star$. This variance may be large for some $\Phi^\star$. One can deal with this problem by preferring a regularized version of this estimator, that is

$$\left(\widetilde{R}_{n,p}^T \widetilde{R}_{n,p} + S_{n,p}\right)^{-1} \widetilde{R}_{n,p}^T \widetilde{\boldsymbol{\rho}}_{n,2:(p+1)}^T , \tag{14}$$

where $S_{n,p}$ is a $p\times p$ positive semi-definite symmetric real matrix. If $S_{n,p} = o_P\left(n^{-1/2}\right)$, then Proposition 2.1 remains true with $\widetilde{\Phi}_n^{(p)}$ being replaced by the expression of Equation (14).

3 Change-points and expectations estimation

In this section, the number of change-points $m^\star$ is assumed to be known. In the sequel, for notational simplicity, $m^\star$ will be denoted by m. Our goal is to estimate both the change-points and the means in Model (1). A first idea consists in maximizing the Gaussian quasi-likelihood conditioned on $y_0,\dots,y_{1-p}$ and to minimize it with respect to Φ. Due to a quadratic term that involves both δ_{k-1} and δ_k, this criterion cannot be efficiently minimized. Therefore, we propose to use an alternative criterion defined as follows:

$$SS_m(y,\Phi,\boldsymbol{\delta},\boldsymbol{t}) = \sum_{k=0}^{m}\sum_{i=t_k+1}^{t_{k+1}}\left(y_i - \sum_{r=1}^{p}\phi_r y_{i-r} - \delta_k\right)^2 .$$

Note that minimizing $SS_m\left(z,\Phi,\left(1-\sum_{r=1}^p\phi_r\right)\boldsymbol{\mu},\boldsymbol{t}\right)$ is equivalent to maximizing the Gaussian log-likelihood, conditioned on $z_0,\dots,z_{1-p}$, of the following model maximized with respect to σ:

$$z_i - \mu_k^\star = \sum_{r=1}^{p}\phi_r\left(z_{i-r} - \mu_k^\star\right) + \epsilon_i , \; t_{n,k}^\star + 1 \le i \le t_{n,k+1}^\star , \; 0\le k\le m , \; 1\le i\le n , \tag{15}$$

where the ϵ_i's are defined as in Model (1). In this model, as in the models considered in Bardet et al. (2012), the expectation changes are not abrupt anymore as in Model (1).

Proposition 3.1. *Let* $\left(\overline{\Phi}_n\right)$ *be a sequence of rv's on* $\mathbb{R}^p$ *and* $z = (z_{1-p},\dots,z_n)$ *a finite sequence of real-valued rv's satisfying* (15). *Let* $\widehat{\boldsymbol{\delta}}_n(z,\overline{\Phi}_n)$ *and* $\widehat{\boldsymbol{t}}_n(z,\overline{\Phi}_n)$ *be defined by*

$$\left(\widehat{\boldsymbol{\delta}}_n(z,\overline{\Phi}_n), \widehat{\boldsymbol{t}}_n(z,\overline{\Phi}_n)\right) = \underset{(\boldsymbol{\delta},\boldsymbol{t})\in\mathbb{R}^{m+1}\times\mathcal{A}_{n,m}}{\arg\min} SS_m\left(z,\overline{\Phi}_n,\boldsymbol{\delta},\boldsymbol{t}\right) , \tag{16}$$

$$\widehat{\boldsymbol{\tau}}_n(z,\overline{\Phi}_n) = \frac{1}{n}\widehat{\boldsymbol{t}}_n(z,\overline{\Phi}_n), \tag{17}$$

where

$$\mathcal{A}_{n,m} = \{(t_0, \ldots, t_{m+1})\,; t_0 = 0 < t_1 < \cdots < t_m < t_{m+1} = n, \forall k = 1, \ldots, m+1, t_k - t_{k-1} \geq \Delta_n\} \qquad (18)$$

and where (Δ_n) *is a real sequence such that* $n^{-\alpha}\Delta_n \longrightarrow \infty$*, as* $n \to \infty$ *and* $\alpha > 0$*. Assume that*

$$\left(\overline{\Phi}_n - \Phi^\star\right) = O_P\left(n^{-1/2}\right), \qquad (19)$$

as n *tends to infinity. Then,*

$$\|\widehat{\boldsymbol{\tau}}_n(z, \overline{\Phi}_n) - \boldsymbol{\tau}^\star\| = O_P\left(n^{-1}\right), \qquad \|\widehat{\boldsymbol{\delta}}_n(z, \overline{\Phi}_n) - \boldsymbol{\delta}^\star\| = O_P\left(n^{-1/2}\right),$$

where $\|\cdot\|$ *is the Euclidian norm.*

Proposition 3.2. *The result of Proposition 3.1 still holds under the same assumptions when* z *is replaced with* y *satisfying* (1)*.*

The proofs of Propositions 3.1 and 3.2 are given in Section 6.2 and 6.3, respectively. Note that the estimators defined in these propositions have the same asymptotic properties as those of the estimators proposed by Lavielle and Moulines (2000). Also, the estimator $\widetilde{\Phi}_n^{(p)}$ defined in Section 2 satisfies the same properties as $\overline{\Phi}_n$ and can thus be used in the criterion SS_m for providing consistent estimators of the change-points and of the means.

4 Selecting the number of change-points

4.1 Criterion to select the number of change-points, the order of the autoregression being fixed

In this section, we propose to adapt the modified Bayesian information criterion (mBIC, Zhang and Siegmund, 2007) to our autoregressive noise framework. mBIC was proposed to select the number m of change-points in the mean in the particular case of segmentation of an independent Gaussian process x. This criterion is derived from an $O_P(1)$ approximation of the Bayes factor between models with m and 0 change-points, respectively. Its performance for model selection is assessed by simulation studies (Zhang, 2005; Frick et al., 2014).

The mBIC selection procedure consists in choosing the number of change-points as:

$$\widehat{m} = \arg\max_m C_m(x, 0)\ , \qquad (20)$$

where the criterion $C_m(y, \Phi)$ is defined for a process y as

$$C_m(y, \Phi) = \\ -\left(\frac{n-m+1}{2}\right)\log SS_m(y,\Phi) + \log\Gamma\left(\frac{n-m+1}{2}\right) - \frac{1}{2}\sum_{k=0}^{m}\log n_k(\widehat{\boldsymbol{t}}(y,\Phi)) - m\log n. \qquad (21)$$

In the latter equation

$$SS_m(y, \Phi) = \min_{\boldsymbol{\delta}, \boldsymbol{t}} SS_m(y, \Phi, \boldsymbol{\delta}, \boldsymbol{t}), \qquad (22)$$

where the minimization with respect to $\boldsymbol{t}$ is performed in $\mathcal{A}_{n,m}$ defined in (18), and

$$n_k(\widehat{\boldsymbol{t}}(y, \Phi)) = \widehat{t}_{k+1}(y, \Phi) - \widehat{t}_k(y, \Phi),$$

where $\widehat{\boldsymbol{t}}(y,\Phi) = (\widehat{t}_1(y,\Phi),\ldots,\widehat{t}_m(y,\Phi))$ is defined as $\widehat{\boldsymbol{t}}(y,\Phi) = \underset{\boldsymbol{t}\in\mathcal{A}_{n,m}}{\arg\min}\min_{\delta} SS_m(y,\Phi,\delta,\boldsymbol{t})$.

Note that, in Model (15), the criterion could be directly applied to the decorrelated series $v^\star = (v_i^\star)_{1\le i\le n} = (y_i - \sum_{r=1}^{p}\phi_r^\star y_{i-r})_{1\le i\le n}$ since

$$C_m(y,\Phi^\star) = C_m(v^\star,0).$$

We propose to use the same selection criterion, replacing $\Phi^\star$ by some relevant estimator $\overline{\Phi}_n$. The following two propositions show that this plug-in approach results in the same asymptotic properties under both Models (15) and (1).

Proposition 4.1. *For any positive m, for a process z satisfying* (15) *and under the assumptions of Proposition 3.1, we have*

$$C_m(z,\overline{\Phi}_n) \;=\; C_m(z,\Phi^\star) + O_P(1),\;\; \text{as } n\to\infty\,.$$

Proposition 4.2. *For any positive m, for a process y satisfying* (1) *and under the assumptions of Proposition 3.2 , we have*

$$C_m(y,\overline{\Phi}_n) \;=\; C_m(y,\Phi^\star) + O_P(1),\;\; \text{as } n\to\infty\,.$$

The proofs of Propositions 4.1 and 4.2 are similar to Chakar et al. (2014, Propositions 6 and 7).

In practice, we propose to take $\overline{\Phi}_n = \widetilde{\Phi}_n^{(p)}$ as defined in Section 2, which satisfies the conditions of Proposition 4.2 to estimate the number of segments by

$$\widehat{m} \;=\; \underset{0\le m\le m_{\max}}{\arg\max}\, C_m\left(y,\widetilde{\Phi}_n^{(p)}\right)\,, \tag{23}$$

where C_m is defined in (21), and for a given maximum number of changes $m_{\max}$.

4.2 Heuristic to select both the number of changes and the autoregression order

Applying Zhang (2005, Theorem 2.2) to the series $v^\star = (v_i^\star)_{1\le i\le n} = (z_i - \sum_{r=1}^{p}\phi_r^\star z_{i-r})_{1\le i\le n}$, for a process z satisfying (15) and with the corresponding priors, gives:

$$\log P(m|v^\star) = C_m(v^\star,0) - C_0(v^\star,0) + \log P(m=0|v^\star) + O_P(1)\,,$$

that is

$$P(m|z,p,\Phi) = C_m(z,\Phi) - C_0(v^\star,0) + \log P(m=0|v^\star) + O_P(1)\,,$$

From Schwarz (1978), we approximate then $\log P(m,p|z)$, up to a constant, by

$$\log P\left(m\,\middle|\,z,p,\widehat{\Phi}_{ML}^{(p)}\right) - \frac{p}{2}\log n\,,$$

and then, up to a constant, by

$$C_m\left(z,\widehat{\Phi}_{ML}^{(p)}\right) - \frac{p}{2}\log n\,,$$

where $\widehat{\Phi}_{ML}^{(p)}$ is the maximum likelihood estimator of $\Phi^\star$ in the model with m changes and autoregression order p. Replacing $\widehat{\Phi}_{ML}^{(p)}$ by $\widetilde{\Phi}_n^{(p)}$, and z by y satisfying (1), we propose to select m and p by

$$(\widehat{m}',\widehat{p}') = \underset{0\le m\le m_{\max},0\le p\le p_{\max}}{\arg\max}\left\{C_m\left(y,\widetilde{\Phi}_n^{(p)}\right) - \frac{p}{2}\log n\right\}\,, \tag{24}$$

with given $m_{\max}$ and $p_{\max}$. Even if we do not aim to estimate the order p of the autoregression, this criterion is interesting by being more flexible than (23). Indeed, if at the true order p, $\widetilde{\Phi}_n^{(p)}$ provides a poor estimate of $\Phi^\star$, a different order (e.g. $p+1$) can lead to a better fitting of the model and then a better estimate of the number of changes.

5 Numerical experiments

5.1 Practical implementation

Our decorrelation procedure introduces spurious change-points in the series, at distance $\leq p$ of the true change-points. When the length of the series tends to infinity, the effect of these artefacts on estimates vanishes, but with a finite length series our procedure may be affected. To fix this point, we propose a post-processing to the estimated change-points $\widehat{\boldsymbol{t}}_{\boldsymbol{n}}$, which consists in removing segments of length smaller than p:

$$PP\left(\widehat{\boldsymbol{t}}_n, p\right) = \left\{\widehat{t}_{n,k} \in \widehat{\boldsymbol{t}}_n\right\} \smallsetminus \left\{\widehat{t}_{n,i} \textit{ such that } \exists j, \widehat{t}_{n,i} - p \leq \widehat{t}_{n,j} < \widehat{t}_{n,i} \textit{ and } \left(j = 1 \textit{ or } \widehat{t}_{n,j-1} < \widehat{t}_{n,j} - p\right)\right\} .$$

5.2 Simulation design

We simulated Gaussian $AR(p)$ series $y_1, \ldots, y_n$ of length $n = 7200$ and $n = 14400$, with $p = 2$ or $p = 5$. Additionally, the observations $y_{-19}, \ldots, y_0$, are simulated and used for conditioning the quasi-likelihood. All series were affected by $m^\star = 6$ change-points located at fractions $1/6 \pm 1/36, 3/6 \pm 2/36, 5/6 \pm 3/36$ of their length. The mean within each segment alternates between 0 and 1, starting with $\mu_1 = 0$. We considered autoregression parameters that verify the assumptions to get a stationary causal process, see Brockwell and Davis (1991, Theorem 3.1.1). We focused our attention on some parameters for which the computed estimators have a typical behavior and are interesting to illustrate our method.

For $p = 2$ the parameters are the following

$$(\phi_1^\star, \phi_2^\star, \sigma^\star) \in \{(-1.2, -0.4, 0.4), (1.6, -0.8, 0.4), (0.2, 0.2, 0.4), (0.2, 0.6, 0.4), (0.4, 0.2, 0.2)\} .$$

For $p = 5$ the parameters are the following

$$(\phi_1^\star, \phi_2^\star, \phi_3^\star, \phi_4^\star, \phi_5^\star, \sigma^\star) \in \{(0.5, 0, 0, 0.5, -0.5, 0.4), (0.5, 0, 0, 0, -0.5, 0.4)\} .$$

Each combination was replicated $S = 100$ times.

5.3 Quality criteria

To assess the quality of the estimation of the autoregression parameters, we used the Root-Mean-Square Errors (RMSE) of $\widehat{\Phi}_n^{(p)}$ defined in (5).

To study the performances of our proposed model selection criteria, we computed and compared:

- $\widehat{m}^0 = \arg\max_m C_m(y, 0)$, the criterion C_m without any decorrelation procedure,

- $\widehat{m}_Y^0$, the estimated number of changes derived from the BIC-type penalized criterion defined by Yao (1988),
- $\widehat{m}^\star = \arg\max_m C_m(y, \Phi^\star)$, the criterion C_m where the series is exactly decorrelated,
- $\widehat{m}_{PP}^\star$, the post-processed number of changes, that is the number of changes of $PP(\widehat{\boldsymbol{t}}_n, p)$ if $\widehat{\boldsymbol{t}}_n$ is the vector of the estimated change-point locations obtained by the minimization (22) with $\Phi = \Phi^\star$,
- $\widehat{m} = \arg\max_m C_m(y, \widehat{\Phi}_n^{(p)})$, and the result of post-processing $\widehat{m}_{PP}$,
- $(\widehat{m}', \widehat{p}')$ defined in (24). The post-processing, giving $\widehat{m}'_{PP}$, is performed with $PP(\cdot, \widehat{p}')$,

where C_m is defined in (21). We were particularly interested in the comparison of:

- $\widehat{m}_Y^0$ and $\widehat{m}^0$ to the other estimates to illustrate how much our method improves the estimation of the number of changes,
- $\widehat{m}^\star$ and $\widehat{m}_{PP}^\star$ are compared to $\widehat{m}$ and $\widehat{m}_{PP}$ to identify the errors coming from the estimation of $\Phi^\star$ by $\widetilde{\Phi}_n$,
- the post-processed estimates are compared to the non post-processed estimates to assess the usefulness of this finite-sample correction of our method.

In order to measure the performance of change-point location estimators, we plotted their frequencies. In particular we are interested in the following change-point estimates:

- $\widehat{t}_n^0$, the minimizer on $\mathcal{A}_{n,\widehat{m}^0}$ of $\min_{\boldsymbol{\delta}} SS_{\widehat{m}^0}(\boldsymbol{y}, 0, \boldsymbol{\delta}, \cdot)$.
- $\widehat{t}_{n,PP} = PP\left(\widehat{t}_n, p\right)$, where $\widehat{t}_n$ minimizes $\min_{\boldsymbol{\delta}} SS_{\widehat{m}}\left(\boldsymbol{y}, \widetilde{\Phi}_n^{(p)}, \boldsymbol{\delta}, \cdot\right)$ on $\mathcal{A}_{n,\widehat{m}}$.
- $\widehat{t}'_{n,PP} = PP\left(\widehat{t}'_n, p'\right)$, where $\widehat{t}'_n$ minimizes $\min_{\boldsymbol{\delta}} SS_{\widehat{m}'}\left(\boldsymbol{y}, \widetilde{\Phi}_n^{(p')}, \boldsymbol{\delta}, \cdot\right)$ on $\mathcal{A}_{n,\widehat{m}'}$.

To highlight the peaks, we plotted also the Gaussian kernel density estimate. The bandwith is selected following the method of Sheather and Jones (1991).

5.4 Results

For $p = 2$ the simulation results suggest that the decorrelation procedure is not necessary in all cases. If $\phi_1^\star$ and $\phi_2^\star$ are negative, $\widehat{m}^0 = 6$ almost all the times (see Table 1 and Figures 1 and 2). If only one of the parameters is negative, problems can arise without decorrelation if the other parameter is positively large (see Table 2 and Figures 3 and 4). The core of the problems with $\widehat{m}^0$ is located at $\phi_1^\star > 0, \phi_2^\star > 0$. In these cases, our decorrelation procedure is required (see Tables 3, 4, and 5, and Figures 5, 6, 7, 8, 9, and 10). $\widehat{m}$ can provide poor estimates of m because of a poor preliminary estimate of $\Phi^\star$. However, in this case, $\widehat{m}'$ can provide better estimates, thanks to an overestimation of p by $\widehat{p}'$, as noticed in Section 4.2 (see Tables 4, 5, and Figures 6, 7, 8, 9, 10).

Tables 4 and 5 illustrate the usefulness of post-processing.

For $p = 5$ we can see from Tables 6 and 7 that our method, when p is known, has the same performance as the methodology which would have access to the autoregression parameters $\Phi^\star$, see the lines $\widehat{m}$ and $\widehat{m}^\star$. These performances are not altered by the choice of p, even if it is overestimated by our model selection approach, see the lines $\widehat{m}'$ and $\widehat{p}'$. Moreover, our method outperforms a methodology which would ignore the dependence structure of the process, see the line $\widehat{m}^0$ of these tables. In addition, our method is not only able to select the true number of change-points, whatever p, but also the true change-point positions, as displayed in Figures 12 and 14.

6 Proofs

6.1 Proof of Proposition 2.1

Since there is only a finite number of atypical values in the process (x_i), Theorem 4 of Lévy-Leduc et al. (2011) still holds and gives that for all fixed $h \geq 1$:

$$\sqrt{n-h}\left(\frac{Q_n^2(x_h^+) - Q_n^2(x_h^-)}{4} - \gamma(h)\right) = \frac{1}{\sqrt{n-h}}\sum_{i=1}^{n-h}\psi(\nu_i, \nu_{i+h}) + o_P(1) ,$$

where γ denotes the autocovariance of (ν_i) and where for all x and y,

$$\psi : (x, y) \mapsto \left\{(\gamma(0)+\gamma(h))\, \mathrm{IF}\left(\frac{x+y}{\sqrt{2(\gamma(0)+\gamma(h))}}, Q, \Phi\right) - (\gamma(0)-\gamma(h))\, \mathrm{IF}\left(\frac{x-y}{\sqrt{2(\gamma(0)-\gamma(h))}}, Q, \Phi\right)\right\} .$$

In this equation IF is defined by

$$\mathrm{IF}(x, Q, \Phi) = c(\Phi)\left(\frac{1/4 - \Phi(x + 1/c(\Phi)) + \Phi(x - 1/c(\Phi))}{\int_{\mathbb{R}} \phi(y)\phi(y + 1/c(\Phi))\mathrm{d}y}\right) ,$$

where $c(\Phi) = 1/(\sqrt{2}\Phi^{-1}(5/8)) \approx 2.21914$, and Φ is here the cumulative distribution function of the standard normal distribution. By Theorem 2 of Lévy-Leduc et al. (2011), we obtain that

$$\frac{Q_n^2(x_h^+) + Q_n^2(x_h^-)}{4} - \gamma(0) = o_p(1) .$$

Let $\widehat{\gamma}(0) = (Q_n^2(x_h^+) + Q_n^2(x_h^-))/4$ then

$$\sqrt{n-h}\,(\widehat{\rho}(h) - \rho(h)) = \frac{\widehat{\gamma}(0)^{-1}}{\sqrt{n-h}}\sum_{i=1}^{n-h}\psi(\nu_i, \nu_{i+h}) + o_P(1) ,$$

In order to prove a central limit theorem for $\widehat{\boldsymbol{\rho}}_{1:(p+1)}$, it is enough to prove by the Cramér-Wold device (Billingsley, 1995, Theorem 29.4), that for any a_k in $\mathbb{R}$:

$$\frac{\widehat{\gamma}(0)^{-1}}{\sqrt{n}}\sum_{i=1}^{n-(p+1)}\sum_{k=1}^{p+1} a_k \psi(\nu_i, \nu_{i+k})$$

converges in distribution to a centered Gaussian rv. By Lemma 13 of Lévy-Leduc et al. (2011), ψ is of Hermite rank 2. Hence, the Hermite rank of the linear combination of the ψ function is larger than 2. Thus, Condition (2.40) of Theorem 4 in Arcones (1994) is satisfied and the quantity

$n^{-1/2}\sum_{i=1}^{n-(p+1)}\sum_{k=1}^{p+1} a_k\psi(\nu_i,\nu_{i+k})$ converges in distribution to a centered Gaussian rv with variance $\widetilde{\sigma}^2$ given by

$$\widetilde{\sigma}^2 = \mathbb{E}\left[\left(\sum_{k=1}^{p+1} a_k\psi(\nu_1,\nu_{k+1})\right)^2\right] + 2\sum_{\ell\geq 1}\mathbb{E}\left[\left(\sum_{k=1}^{p+1} a_k\psi(\nu_1,\nu_{k+1})\right)\left(\sum_{k=1}^{p+1} a_k\psi(\nu_{\ell+1},\nu_{k+\ell+1})\right)\right] .$$

By Slutsky's lemma, $\sqrt{n-h}\,(\widehat{\rho}(h)-\rho(h))$ converges in distribution to a centered Gaussian rv with variance $\gamma(0)\widetilde{\sigma}^2$. Since $(x_i - \mathbb{E}x_i)$ is an ARMA(p,1) process, with autoregressive parameters $\phi_1^\star,\ldots,\phi_p^\star$, we get, by Azencott and Dacunha-Castelle (1986, Chapter 11, Paragraph 2), that R_p, as defined in (9), is invertible, and

$$\Phi^\star = R_p^{-1}\boldsymbol{\rho}_{2:(p+1)}^T ,$$

where $\boldsymbol{\rho}_{2:(p+1)}$ is defined in (7).

Let $g:\mathcal{D}\subset\mathbb{R}^{p+1}\to\mathbb{R}^p$ defined by

$$g(u) = \left(u_{|j-i-1|}\mathbf{1}_{j-i-1\neq 0} + \mathbf{1}_{j-i-1=0}\right)_{1\leq i,j\leq p}^{-1}\left(u_2,\ldots,u_{p+1}\right)^T . \tag{25}$$

Hence,

$$\begin{aligned}\Phi^\star &= g(\boldsymbol{\rho}_{1:(p+1)}) ,\\ \widetilde{\Phi}_n^{(p)} &= g(\widetilde{\boldsymbol{\rho}}_{\boldsymbol{n}\,1:(p+1)}) .\end{aligned}$$

By (11) and the delta method:

$$\sqrt{n}\left(\widetilde{\Phi}_n^{(p)} - \Phi^\star\right) \underset{n\to\infty}{\to} \mathcal{N}\left(0,\nabla g(\boldsymbol{\rho}_{1:(p+1)})^T V\nabla g(\boldsymbol{\rho}_{1:(p+1)})\right) ,$$

$\nabla g(\boldsymbol{\rho}_{1:(p+1)})$ being the Jacobian matrix of g in $\boldsymbol{\rho}_{1:(p+1)}$. Let us determine $\nabla g(u)\cdot h$ for all $u\in\mathcal{D}, h\in\mathbb{R}^{p+1}$. Using (25), we get

$$\left(h_{|j-i-1|}\mathbf{1}_{j-i-1\neq 0}\right)_{1\leq i,j\leq p} g(u) + \left(u_{|j-i-1|}\mathbf{1}_{j-i-1\neq 0} + \mathbf{1}_{j-i-1=0}\right)_{1\leq i,j\leq p}\nabla g(u)\cdot h = \left(h_2,\ldots,h_{p+1}\right)^T .$$

Applied to $u=\boldsymbol{\rho}_{1:(p+1)}$, we have

$$R_p\nabla g\left(\boldsymbol{\rho}_{1:(p+1)}\right)\cdot h = \left(h_{j+1} - \sum_{j=1}^{i}\phi_{i+1-j}^\star h_j + \sum_{j=1}^{p-i-1}\phi_{i+j+1}^\star h_j\right)_{1\leq j\leq p} ,$$

and then $\nabla g\left(\boldsymbol{\rho}_{1:(p+1)}\right) = R_p^{-1}M$, where M is defined in (13).

6.2 Proof of Proposition 3.1

In the sequel, we need the following definitions, notations and remarks. Observe that (15) can be rewritten as follows:

$$z = \sum_{r=1}^{p}\phi_r^\star B^r z + T\left(\boldsymbol{t}_n^\star\right)\boldsymbol{\delta}^\star + \epsilon , \tag{26}$$

where

$$z = \begin{pmatrix} z_1 \\ \vdots \\ z_n\end{pmatrix}, \quad B^r z = \begin{pmatrix} z_{1-r} \\ \vdots \\ z_{n-r}\end{pmatrix}, \quad \boldsymbol{\delta}^\star = \begin{pmatrix} \delta_0^\star \\ \vdots \\ \delta_m^\star\end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n\end{pmatrix}, \tag{27}$$

where $\delta_k^\star = (1-\sum_{r=1}^p\phi_r^\star)\mu_k^\star$, for $0\leq k\leq m$, and $T(\boldsymbol{t})$ is an $n\times(m+1)$ matrix where the kth column is $(\underbrace{0,\ldots,0}_{t_{k-1}}\underbrace{1,\ldots,1}_{t_k-t_{k-1}}\underbrace{0,\ldots,0}_{n-t_k})^T$.

Let us define the exact and estimated decorrelated series by

$$w^\star = z - \sum_{r=1}^{p} \phi_r^\star B^r z\,, \tag{28}$$

$$\overline{w} = z - \sum_{r=1}^{p} \overline{\phi}_{r,n} B^r z\,. \tag{29}$$

where $\overline{\Phi}_n = (\phi_{r,n})_{1\le r\le p}$.

For any vector subspace E of $\mathbb{R}^n$, let π_E denote the orthogonal projection of $\mathbb{R}^n$ on E. Let also $\|\cdot\|$ be the Euclidean norm on $\mathbb{R}^n$, $\langle\cdot,\cdot\rangle$ the canonical scalar product on $\mathbb{R}^n$ and $\|\cdot\|_\infty$ the sup norm.

For x a vector of $\mathbb{R}^n$ and $\boldsymbol{t}\in\mathcal{A}_{n,m}$, let

$$J_{n,m}(x,\boldsymbol{t}) = \frac{1}{n}\left(\|\pi_{E_{\boldsymbol{t}_n^\star}}(x)\|^2 - \|\pi_{E_{\boldsymbol{t}}}(x)\|^2\right)\,, \tag{30}$$

written $J_n(x,\boldsymbol{t})$ in the sequel for notational simplicity. In (30), $E_{\boldsymbol{t}_n^\star}$ and $E_{\boldsymbol{t}}$ correspond to the linear subspaces of $\mathbb{R}^n$ generated by the columns of $T(\boldsymbol{t}_n^\star)$ and $T(\boldsymbol{t})$, respectively. We shall use the same decomposition as the one introduced in Lavielle and Moulines (2000):

$$J_n(x,\boldsymbol{t}) = K_n(x,\boldsymbol{t}) + V_n(x,\boldsymbol{t}) + W_n(x,\boldsymbol{t})\,, \tag{31}$$

where

$$\begin{aligned}
K_n(x,\boldsymbol{t}) &= \frac{1}{n}\left\|\left(\pi_{E_{\boldsymbol{t}_n^\star}} - \pi_{E_{\boldsymbol{t}}}\right)\mathbb{E}x\right\|^2\,,\\
V_n(x,\boldsymbol{t}) &= \frac{1}{n}\left(\left\|\pi_{E_{\boldsymbol{t}_n^\star}}(x-\mathbb{E}x)\right\|^2 - \|\pi_{E_{\boldsymbol{t}}}(x-\mathbb{E}x)\|^2\right)\,,\\
W_n(x,\boldsymbol{t}) &= \frac{2}{n}\left(\left\langle\pi_{E_{\boldsymbol{t}_n^\star}}(x-\mathbb{E}x), \pi_{E_{\boldsymbol{t}_n^\star}}(\mathbb{E}x)\right\rangle - \langle\pi_{E_{\boldsymbol{t}}}(x-\mathbb{E}x), \pi_{E_{\boldsymbol{t}}}(\mathbb{E}x)\rangle\right)\,.
\end{aligned}$$

We shall also use the following notations:

$$\underline{\lambda} = \min_{1\le k\le m} |\delta_k^\star - \delta_{k-1}^\star|\,, \tag{32}$$

$$\overline{\lambda} = \max_{1\le k\le m} |\delta_k^\star - \delta_{k-1}^\star|\,, \tag{33}$$

$$\Delta_{\boldsymbol{\tau}^\star} = \min_{1\le k\le m+1} (\tau_k^\star - \tau_{k-1}^\star)\,, \tag{34}$$

$$\mathcal{C}_{\alpha,\gamma,n} = \left\{\boldsymbol{t}\in\mathcal{A}_{n,m};\, \alpha\underline{\lambda}^{-2} \le \|\boldsymbol{t}-\boldsymbol{t}_n^\star\| \le n\gamma\Delta_{\boldsymbol{\tau}^\star}\right\}\,, \tag{35}$$

$$\mathcal{C}'_{\alpha,\gamma,n} = \mathcal{C}_{\alpha,\gamma,n} \cap \left\{\boldsymbol{t}\in\mathcal{A}_{n,m};\, \forall k=1,\dots,m, t_k \ge t_{n,k}^\star\right\}\,, \tag{36}$$

$$\begin{aligned}\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I}) = \Big\{&\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n};\\
&\forall k\in\mathcal{I}, \alpha\underline{\lambda}^{-2} \le t_k - t_{n,k}^\star \le n\gamma\Delta_{\boldsymbol{\tau}^\star} \text{ and } \forall k\notin\mathcal{I}, t_k - t_{n,k}^\star < \alpha\underline{\lambda}^{-2}\Big\}\,,\end{aligned} \tag{37}$$

for any $\alpha>0$, $0<\gamma<1/2$ and $\mathcal{I}\subset\{1,\dots,m\}$. We shall also need the following lemmas in order to prove Proposition 3.1 which are proved below.

Lemma 6.1. *Let $(z_{1-p},\dots,z_n)$ be defined by (1) or (15), then, for all $r=0,\dots,p$:*

$$\|B^r z\| = O_P\left(n^{1/2}\right)\,,$$

as n tends to infinity, where $B^r z$ is defined in (27).

Lemma 6.2. *Let $(z_{1-p},\dots,z_n)$ be defined by (1) or (15) then, for all $\boldsymbol{t}\in\mathcal{A}_{n,m}$,*

$$|J_n(\overline{w},\boldsymbol{t}) - J_n(w^\star,\boldsymbol{t})| \le \frac{2}{n}\sum_{r=1}^{p}\left|\phi_r^\star - \overline{\phi}_{r,n}\right|\|B^r z\|\left(p\left|\phi_r^\star - \overline{\phi}_{r,n}\right|\|B^r z\| + 2\|w^\star\|\right) = O_P\left(n^{-1/2}\right)\,,$$

where J_n is defined in (30), Bz and z are defined in (27), $w^\star$ is defined in (28) and $\overline{w}$ is defined in (29).

Lemma 6.3. *Under the assumptions of Proposition 3.1, $\|\overline{\boldsymbol{\tau}}_{\boldsymbol{n}} - \boldsymbol{\tau}^\star\|_\infty$ converges in probability to 0, as n tends to infinity.*

Lemma 6.4. *Under the assumptions of Proposition 3.1 and for any $\alpha > 0$, $0 < \gamma < 1/2$ and $\mathcal{I} \subset \{1,\dots,m\}$,*

$$P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})}\left(\frac{1}{2}K_n\left(w^\star,\boldsymbol{t}\right) + V_n\left(w^\star,\boldsymbol{t}\right) + W_n\left(w^\star,\boldsymbol{t}\right)\right) \le 0\right) \longrightarrow 0\ ,\ \textit{as}\ n \to \infty\ ,$$

where $\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})$ is defined in (37) *and $w^\star$ is defined in* (28).

Lemma 6.5. *Under the assumptions of Proposition 3.1 and for any $\alpha > 0$, $0 < \gamma < 1/2$ and $\mathcal{I} \subset \{1,\dots,m\}$,*

$$P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})} J_n\left(\overline{w},\boldsymbol{t}\right) \le 0\right) \longrightarrow 0\ ,\ \textit{as}\ n \to \infty\ ,$$

where $\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})$ is defined in (37) *and $\overline{w}$ is defined in* (29).

Lemma 6.6. *Under the assumptions of Proposition 3.1,*

$$\|\widehat{\boldsymbol{\tau}}_{\boldsymbol{n}}(z,\overline{\Phi}_n) - \boldsymbol{\tau}^\star\|_\infty = O_P\left(n^{-1}\right)\ .$$

Proof of Lemma 6.1. Without loss of generality, assume $(z_{1-p},\dots,z_n)$ is defined by (15). $\|B^r z\|^2 = \sum_{i=1-r}^{n-r} z_i^2$ then Markov inequality implies that $\|B^r z\|^2 = O_P(n)$. □

Proof of Lemma 6.2. By (28), $\overline{w} = w^\star + \sum_{r=1}^p(\phi_r^\star - \overline{\phi}_{r,n})B^r z$. We get

$$\begin{aligned}
\|\pi_{E_{\boldsymbol{t}}}(\overline{w})\|^2 - \|\pi_{E_{\boldsymbol{t}}}(w^\star)\|^2 &= \left\|\pi_{E_{\boldsymbol{t}}}(w^\star) + \sum_{r=1}^p(\phi_r^\star - \overline{\phi}_{r,n})\pi_{E_{\boldsymbol{t}}}(B^r z)\right\|^2 - \|\pi_{E_{\boldsymbol{t}}}(w^\star)\|^2\\
&= \left\|\sum_{r=1}^p(\phi_r^\star - \overline{\phi}_{r,n})\pi_{E_{\boldsymbol{t}}}(B^r z)\right\|^2 + 2\sum_{r=1}^p(\phi_r^\star - \overline{\phi}_{r,n})\left\langle\pi_{E_{\boldsymbol{t}}}(w^\star),\pi_{E_{\boldsymbol{t}}}(B^r z)\right\rangle\\
&\le \sum_{r=1}^p p(\phi_r^\star - \overline{\phi}_{r,n})^2\,\|\pi_{E_{\boldsymbol{t}}}(B^r z)\|^2 + 2\sum_{r=1}^p(\phi_r^\star - \overline{\phi}_{r,n})\left\langle\pi_{E_{\boldsymbol{t}}}(w^\star),\pi_{E_{\boldsymbol{t}}}(B^r z)\right\rangle\\
&\le \sum_{r=1}^p(\phi_r^\star - \overline{\phi}_{r,n})\left(p(\phi_r^\star - \overline{\phi}_{r,n})\,\|\pi_{E_{\boldsymbol{t}}}(B^r z)\|^2 + 2\left\langle\pi_{E_{\boldsymbol{t}}}(w^\star),\pi_{E_{\boldsymbol{t}}}(B^r z)\right\rangle\right)\\
&\le \sum_{r=1}^p(\phi_r^\star - \overline{\phi}_{r,n})\left\langle\pi_{E_{\boldsymbol{t}}}(B^r z), p(\phi_r^\star - \overline{\phi}_{r,n})\pi_{E_{\boldsymbol{t}}}(B^r z) + 2\pi_{E_{\boldsymbol{t}}}(w^\star)\right\rangle\ .
\end{aligned}$$

The Cauchy-Schwarz inequality and the 1-Lipschitz property of projections give

$$\left|\|\pi_{E_{\boldsymbol{t}}}(\overline{w})\|^2 - \|\pi_{E_{\boldsymbol{t}}}(w^\star)\|^2\right| \le \sum_{r=1}^p\left|\phi_r^\star - \overline{\phi}_{r,n}\right|\|B^r z\|\left(p\left|\phi_r^\star - \overline{\phi}_{r,n}\right|\|B^r z\| + 2\|w^\star\|\right)$$

The conclusion follows from (30), (19) and Lemma 6.1. □

Proof of Lemma 6.3. Lavielle and Moulines (2000, proof of Theorem 3) give the following bounds for any $\boldsymbol{t} \in \mathcal{A}_{n,m}$:

$$K_n\left(w^\star,\boldsymbol{t}\right) \ \ge\ \underline{\lambda}^2\min\left(\frac{1}{n}\max_{1\le k\le m}\left|t_k - t_{n,k}^\star\right|, \Delta_{\boldsymbol{\tau}^\star}\right)\ , \tag{38}$$

$$V_n\left(w^\star,\boldsymbol{t}\right) \ \ge\ -\frac{2\left(m+1\right)}{n\Delta_n}\left(\max_{1\le s\le n}\left(\sum_{i=1}^s\epsilon_i\right)^2 + \max_{1\le s\le n}\left(\sum_{i=n-s}^n\epsilon_i\right)^2\right)\ , \tag{39}$$

$$\left|W_n\left(w^\star,\boldsymbol{t}\right)\right| \ \le\ \frac{3\left(m+1\right)^2\overline{\lambda}}{n}\left(\max_{1\le s\le n}\left|\sum_{i=1}^s\epsilon_i\right| + \max_{1\le s\le n}\left|\sum_{i=n-s}^n\epsilon_i\right|\right)\ , \tag{40}$$

where $\underline{\lambda}$, $\overline{\lambda}$ anf $\Delta_{\boldsymbol{\tau}^\star}$ are defined in (32–34). For any $\alpha > 0$, define, as in the proof of Theorem 3 of Lavielle and Moulines (2000),

$$\mathcal{C}_{n,\alpha} = \{\boldsymbol{t} \in \mathcal{A}_{n,m};\, \|\boldsymbol{t} - \boldsymbol{t}_n^\star\|_\infty \geq n\alpha\} \ . \tag{41}$$

For $0 < \alpha < \Delta_{\boldsymbol{\tau}^\star}$, we have:

$$\begin{aligned}
P\left(\left\|\widehat{\boldsymbol{t}}_n(z, \overline{\Phi}_n) - \boldsymbol{t}_n^\star\right\|_\infty \geq n\alpha\right) &\leq P\left(\min_{\boldsymbol{t}\in\mathcal{C}_{n,\alpha}} J_n\left(\overline{w}, \boldsymbol{t}\right) \leq 0\right) \\
&\leq P\left(\min_{\boldsymbol{t}\in\mathcal{C}_{n,\alpha}} \left(J_n\left(\overline{w}, \boldsymbol{t}\right) - J_n\left(w^\star, \boldsymbol{t}\right)\right) \leq -\alpha\underline{\lambda}^2\right) \\
&+ P\left(\min_{\boldsymbol{t}\in\mathcal{C}_{n,\alpha}} \left(V_n\left(w^\star, \boldsymbol{t}\right) + W_n\left(w^\star, \boldsymbol{t}\right)\right) \leq -\alpha\underline{\lambda}^2\right) \\
&\leq P\left(\min_{\boldsymbol{t}\in\mathcal{C}_{n,\alpha}} \left(J_n\left(\overline{w}, \boldsymbol{t}\right) - J_n\left(w^\star, \boldsymbol{t}\right)\right) \leq -\alpha\underline{\lambda}^2\right) \\
&+ P\left(\max_{1\leq s\leq n}\left(\sum_{i=1}^{s}\epsilon_i\right)^2 + \max_{1\leq s\leq n}\left(\sum_{i=n-s}^{n}\epsilon_i\right)^2 \geq c\underline{\lambda}^2 n\Delta_n\alpha\right) \\
&+ P\left(\max_{1\leq s\leq n}\left|\sum_{i=1}^{s}\epsilon_i\right| + \max_{1\leq s\leq n}\left|\sum_{i=n-s}^{n}\epsilon_i\right| \geq c\underline{\lambda}^2 n\alpha\overline{\lambda}^{-1}\right)
\end{aligned}$$

for some positive constant c. The last two terms of this sum go to 0 when n goes to infinity (Lavielle and Moulines, 2000, proof of Theorem 3). To show that the first term shares the same property, it suffices to show that $J_n\left(\overline{w}, \boldsymbol{t}\right) - J_n\left(w^\star, \boldsymbol{t}\right)$ is bounded uniformly in $\boldsymbol{t}$ by a sequence of rv's which converges to 0 in probability. This result holds by Lemma 6.2. □

Proof of Lemma 6.4. Using Lavielle and Moulines (2000, Equations (64–66)), one can show the bound (73) of Lavielle and Moulines (2000) on

$$P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})}\left(K_n\left(w^\star, \boldsymbol{t}\right) + V_n\left(w^\star, \boldsymbol{t}\right) + W_n\left(w^\star, \boldsymbol{t}\right)\right) \leq 0\right).$$

Using the same arguments, we have the same bound on

$$P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})}\left(\frac{1}{2}K_n\left(w^\star, \boldsymbol{t}\right) + V_n\left(w^\star, \boldsymbol{t}\right) + W_n\left(w^\star, \boldsymbol{t}\right)\right) \leq 0\right).$$

We conclude using Lavielle and Moulines (2000, Equations (67–71)). □

Proof of Lemma 6.5. By (31),

$$\begin{aligned}
P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})} J_n\left(\overline{w}, \boldsymbol{t}\right) \leq 0\right) &\leq P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})}\left(J_n\left(\overline{w}, \boldsymbol{t}\right) - J_n\left(w^\star, \boldsymbol{t}\right) + \frac{1}{2}K_n\left(w^\star, \boldsymbol{t}\right)\right) \leq 0\right) \\
&+ P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})}\left(\frac{1}{2}K_n\left(w^\star, \boldsymbol{t}\right) + V_n\left(w^\star, \boldsymbol{t}\right) + W_n\left(w^\star, \boldsymbol{t}\right)\right) \leq 0\right) .
\end{aligned}$$

By Lemma 6.4, the conclusion thus follows if

$$P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})}\left(J_n\left(\overline{w}, \boldsymbol{t}\right) - J_n\left(w^\star, \boldsymbol{t}\right) + \frac{1}{2}K_n\left(w^\star, \boldsymbol{t}\right)\right) \leq 0\right) \longrightarrow 0\ , \text{ as } n \to \infty\ .$$

Since $\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})} K_n\left(w^\star, \boldsymbol{t}\right) \geq (1-\gamma)\,\Delta_{\boldsymbol{\tau}^\star}\alpha$ (Lavielle and Moulines, 2000, Equation (65)),

$$P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})}\left(J_n\left(\overline{w}, \boldsymbol{t}\right) - J_n\left(w^\star, \boldsymbol{t}\right) + \frac{1}{2}K_n\left(w^\star, \boldsymbol{t}\right)\right) \leq 0\right)$$

$$\leq P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})}\left(J_n\left(\overline{w},\boldsymbol{t}\right)-J_n\left(w^\star,\boldsymbol{t}\right)\right)\leq\frac{1}{2}\left(\gamma-1\right)\Delta_{\boldsymbol{\tau}^\star}\alpha\right)\,,$$

and we conclude by Lemma 6.2. ☐

Proof of Lemma 6.6. For notational simplicity, $\widehat{\boldsymbol{t}}_n(z,\overline{\Phi}_n)$ will be replaced by $\overline{\boldsymbol{t}}_n$ in this proof. Since for any $\alpha>0$,

$$P\left(\|\overline{\boldsymbol{t}}_n-\boldsymbol{t}_n^\star\|_\infty<\alpha\underline{\lambda}^{-2}\right)=P(\|\overline{\boldsymbol{t}}_n-\boldsymbol{t}_n^\star\|_\infty\leq n\gamma\Delta_{\boldsymbol{\tau}^\star})-P(\overline{\boldsymbol{t}}_n\in C_{\alpha,\gamma,n})\,,$$

it is enough, by Lemma 6.3, to prove that

$$P\left(\overline{\boldsymbol{t}}_n\in\mathcal{C}_{\alpha,\gamma,n}\right)\longrightarrow 0\,,\ \text{as } n\to\infty\,,$$

for all $\alpha>0$ and $0<\gamma<1/2$. Since $\mathcal{C}_{\alpha,\gamma,n}=\bigcup_{\mathcal{I}\subset\{1,\ldots,m\}}\mathcal{C}_{\alpha,\gamma,n}\cap\left\{\boldsymbol{t}\in\mathcal{A}_{n,m};\forall k\in\mathcal{I},t_k\geq t_{n,k}^\star\right\}$, we shall only study one set in the union without loss of generality and prove that

$$P\left(\overline{\boldsymbol{t}}_n\in\mathcal{C}'_{\alpha,\gamma,n}\right)\longrightarrow 0\,,\ \text{as } n\to\infty\,,$$

where $\mathcal{C}'_{\alpha,\gamma,n}$ is defined in (36). Since $\mathcal{C}'_{\alpha,\gamma,n}=\bigcup_{\mathcal{I}\subset\{1,\ldots,m\}}\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})$, we shall only study one set in the union without loss of generality and prove that

$$P\left(\overline{\boldsymbol{t}}_n\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})\right)\longrightarrow 0\,,\ \text{as } n\to\infty\,.$$

Since

$$P\left(\overline{\boldsymbol{t}}_n\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})\right)\leq P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})}J_n\left(\overline{w},\boldsymbol{t}\right)\leq 0\right)\,,$$

the proof is complete by Lemma 6.5.

☐

Proof of Proposition 3.1. For notational simplicity, $\widehat{\boldsymbol{\delta}}_n(z,\overline{\Phi}_n)$ will be replaced by $\overline{\boldsymbol{\delta}}_n$ in this proof. By Lemma 6.6, the last result to show is

$$\|\overline{\boldsymbol{\delta}}_{\boldsymbol{n}}-\boldsymbol{\delta}^\star\|=O_P\left(n^{-1/2}\right)\,,$$

that is, for all k, $\overline{\delta}_{n,k}-\delta_k^\star=O_P\left(n^{-1/2}\right)$. By (28) and (29),

$$\overline{\delta}_{n,k}=\frac{1}{\overline{t}_{n,k+1}-\overline{t}_{n,k}}\sum_{i=\overline{t}_{n,k}+1}^{\overline{t}_{n,k+1}}\overline{w}_i=\frac{1}{n\left(\overline{\tau}_{n,k+1}-\overline{\tau}_{n,k}\right)}\left(\sum_{i=\overline{t}_{n,k}+1}^{\overline{t}_{n,k+1}}w_i^\star+\sum_{r=1}^{p}\left(\phi_r^\star-\overline{\phi}_{r,n}\right)\sum_{i=\overline{t}_{n,k}+1}^{\overline{t}_{n,k+1}}z_{i-r}\right)\,.$$

By the Cauchy-Schwarz inequality,

$$\left|\sum_{i=\overline{t}_{n,k}+1}^{\overline{t}_{n,k+1}}z_{i-r}\right|\leq\left(\overline{t}_{n,k+1}-\overline{t}_{n,k}\right)^{1/2}\left(z^2_{\overline{t}_{n,k}+1-r}+\cdots+z^2_{\overline{t}_{n,k+1}-r}\right)^{1/2}\leq n^{1/2}\|Bz\|=O_P\left(n\right)\,,$$

where the last equality comes from Lemma 6.1. Hence by (19) and Lemma 6.6,

$$\begin{aligned}\overline{\delta}_{n,k}&=\frac{1}{n\left(\overline{\tau}_{n,k+1}-\overline{\tau}_{n,k}\right)}\sum_{i=\overline{t}_{n,k}+1}^{\overline{t}_{n,k+1}}w_i^\star+O_P\left(n^{-1/2}\right)\\&=\frac{1}{n\left(\overline{\tau}_{n,k+1}-\overline{\tau}_{n,k}\right)}\left(\sum_{i=\overline{t}_{n,k}+1}^{\overline{t}_{n,k+1}}\mathbb{E}w_i^\star+\sum_{i=\overline{t}_{n,k}+1}^{\overline{t}_{n,k+1}}\epsilon_i\right)+O_P\left(n^{-1/2}\right),\end{aligned}$$

where the last equality comes from (26) and (28).

Let us now prove that

$$\frac{1}{n\left(\overline{\tau}_{n,k+1}-\overline{\tau}_{n,k}\right)}\sum_{i=\overline{t}_{n,k}+1}^{\overline{t}_{n,k+1}}\epsilon_i = O_P\left(n^{-1/2}\right)\ . \tag{42}$$

By Lemma 6.3, $n^{-1}\left(\overline{\tau}_{n,k+1}-\overline{\tau}_{n,k}\right)^{-1} = O_P(n^{-1})$. Moreover,

$$\sum_{i=\overline{t}_{n,k}+1}^{\overline{t}_{n,k+1}}\epsilon_i = \sum_{i=t^\star_{n,k}+1}^{t^\star_{n,k+1}}\epsilon_i \pm \sum_{i=\overline{t}_{n,k}+1}^{t^\star_{n,k}}\epsilon_i \pm \sum_{i=t^\star_{n,k+1}}^{\overline{t}_{n,k+1}+1}\epsilon_i\ . \tag{43}$$

By the Central limit theorem, the first term in the right-hand side of (43) is $O_P(n^{1/2})$. By using the Cauchy-Schwarz inequality, we get that the second term of (43) satisfies: $|\sum_{i=\overline{t}_{n,k}+1}^{t^\star_{n,k}}\epsilon_i| \leq |t^\star_{n,k}-\overline{t}_{n,k}|^{1/2}\left(\sum_{i=1}^n \epsilon_i^2\right)^{1/2} = O_P(1)O_P(n^{1/2}) = O_P(n^{1/2})$, by Lemma 6.6. The same holds for the last term in the right-hand side of (43), which gives (42).

Hence,

$$\begin{aligned}
\overline{\delta}_{n,k}-\delta^\star_k &= \frac{1}{n\left(\overline{\tau}_{n,k+1}-\overline{\tau}_{n,k}\right)}\sum_{i=\overline{t}_{n,k}+1}^{\overline{t}_{n,k+1}}\left(\mathbb{E}w^\star_i-\delta^\star_k\right)+O_P\left(n^{-1/2}\right)\\
&= \frac{1}{n\left(\overline{\tau}_{n,k+1}-\overline{\tau}_{n,k}\right)}\sum_{i\in\left\{\overline{t}_{n,k}+1,\ldots,\overline{t}_{n,k+1}\right\}\smallsetminus\left\{t^\star_{n,k}+1,\ldots,t^\star_{n,k+1}\right\}}\left(\mathbb{E}w^\star_i-\delta^\star_k\right)+O_P\left(n^{-1/2}\right)\ ,
\end{aligned}$$

and then

$$\begin{aligned}
\left|\overline{\delta}_{n,k}-\delta^\star_k\right| &\leq \frac{1}{n\left(\overline{\tau}_{n,k+1}-\overline{\tau}_{n,k}\right)}\sharp\left\{\overline{t}_{n,k}+1,\ldots,\overline{t}_{n,k+1}\right\}\smallsetminus\left\{t^\star_{n,k}+1,\ldots,t^\star_{n,k+1}\right\}\max_{l=0,\ldots,m}|\delta^\star_l-\delta^\star_k|\\
&+\ O_P\left(n^{-1/2}\right)\ .
\end{aligned}$$

We conclude by using Lemma 6.6 to get $\sharp\left\{\overline{t}_{n,k}+1,\ldots,\overline{t}_{n,k+1}\right\}\smallsetminus\left\{t^\star_{n,k}+1,\ldots,t^\star_{n,k+1}\right\} = O_P(1)$ and Lemma 6.3 to get $\left(\overline{\tau}_{n,k+1}-\overline{\tau}_{n,k}\right)^{-1} = O_P(1)$. □

6.3 Proof of Proposition 3.2

The connection between Models (1) and (15) is made by the following lemmas.

Lemma 6.7. *Let $(y_0,\ldots y_n)$ be defined by (1) and let*

$$v^\star_i = y_i-\sum_{r=1}^p \phi^\star_r y_{i-1}, \tag{44}$$

$$\Delta^\star_i = \begin{cases}-\left(\mu^\star_k-\mu^\star_{k-1}\right)\sum_{s=r}^p \phi^\star_s & \text{if } i=t^\star_{n,k}+r \text{ and } 1\leq r\leq p\\ 0, & \text{otherwise,}\end{cases} \tag{45}$$

where the $\mu^\star_k$'s are defined in (1), then the process

$$w^\star_i = v^\star_i+\Delta^\star_i \tag{46}$$

equals $z_i-\sum_{r=1}^p z_{i-r}$ where $(z_{1-p},\ldots,z_n)$ verify (15). Such a process $(z_{1-p},\ldots,z_n)$ can be constructed recursively as

$$\begin{cases} z_i &= y_i \text{ for } 1-p\leq i\leq 0\\ z_i &= w^\star_i+\sum_{r=1}^p \phi^\star_r z_{i-r} \text{ for } i>0.\end{cases} \tag{47}$$

Lemma 6.8. *Let $(y_{1-p},\ldots,y_n)$ be defined by* (1) *and let z be defined by (44– 47). Then*

$$\overline{w}_i = \overline{v}_i + \overline{\Delta}_i \tag{48}$$

where

$$\begin{aligned}
\overline{v}_i &= y_i - \sum_{r=1}^{p} \overline{\phi}_{r,n} y_{i-r} && (49)\\
\overline{w}_i &= z_i - \sum_{r=1}^{p} \overline{\phi}_{r,n} z_{i-r} && (50)\\
\overline{\Delta}_i &= \Delta_i^\star + \sum_{r=1}^{p} \left(\phi_r^\star - \overline{\phi}_{r,n}\right)\left(z_{i-r} - y_{i-r}\right) . && (51)
\end{aligned}$$

Lemma 6.9. *Let $\overline{\Delta} = (\overline{\Delta}_i)_{0\le i\le n}$ as defined in* (51)*. Then $\|\overline{\Delta}\| = O_P(1)$.*

Proof of Lemma 6.7. Let z being defined by (47). Using (46), we get, for all $0 \le k \le m, t^\star_{n,k} < i \le t^\star_{n,k+1}$

$$\begin{aligned}
(z_i - \mu_k^\star) - \sum_{r=1}^{p} \phi_r^\star \left(z_{i-r} - \mu_k^\star\right) &= (y_i - \mu_k^\star) - \sum_{r=1}^{p} \phi_r^\star \left(y_{i-r} - \mu_k^\star\right) + \Delta_i^\star \\
&= (y_i - \mu_k^\star) - \sum_{r=1}^{p} \phi_r^\star \left(y_{i-1} - \left(\mu_{k-1}^\star \mathbf{1}_{r\ge i-t^\star_{n,k}} + \mu_k^\star \mathbf{1}_{r<i-t^\star_{n,k}}\right)\right)
\end{aligned}$$

This expression equals $(y_i - \mathbb{E}(y_i)) - \sum_{r=1}^{p} \phi_r^\star \left(y_{i-r} - \mathbb{E}(y_{i-r})\right) = \eta_i - \sum_{r=1}^{p} \phi_r^\star \eta_{i-r} = \epsilon_i$ by (1) and (2). Then z satisfies (15). □

The proof of Lemma 6.8 is straightforward.

Proof of Lemma 6.9. (51) can be written as

$$\overline{\Delta} = \Delta^\star + \sum_{r=1}^{p} \left(\phi_r^\star - \overline{\phi}_{r,n}\right)\left(B^r y - B^r z\right)$$

where $\Delta^\star = (\Delta_i^\star)_{1\le i\le n}$, $B^r y = (y_{i-r})_{1\le i\le n}$ and $B^r z$ is defined in (27). By the triangle inequality,

$$\|\overline{\Delta}\| \le \|\Delta^\star\| + \sum_{r=1}^{p} \left|\phi_r^\star - \overline{\phi}_{r,n}\right| \left(\|B^r y\| + \|B^r z\|\right) . \tag{52}$$

Since $\|\Delta^\star\|$ is constant in n it is bounded. The conclusion follows from (52), (19) and Lemma 6.1. □

Proof of Proposition 3.2. Let y, z, $\overline{v}$, $\overline{w}$ and $\overline{\Delta}$ be defined in Lemma 6.8.

Using (30) and Lemma 6.8, we get

$$J_n(\overline{v},\boldsymbol{t}) = J_n(\overline{w},\boldsymbol{t}) + J_n\left(\overline{\Delta},\boldsymbol{t}\right) - \frac{2}{n}\left(\left\langle \pi_{E_{\boldsymbol{t}_n^\star}}(\overline{w}), \pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{\Delta}\right)\right\rangle - \left\langle \pi_{E_{\boldsymbol{t}}}(\overline{w}), \pi_{E_{\boldsymbol{t}}}\left(\overline{\Delta}\right)\right\rangle\right).$$

By the Cauchy-Schwarz inequality and the 1-Lipschitz property of projections, we have

$$\begin{aligned}
\left|J_n\left(\overline{\Delta},\boldsymbol{t}\right)\right| &\le 2\|\overline{\Delta}\|^2,\\
\left|\left\langle \pi_{E_{\boldsymbol{t}_n^\star}}(\overline{w}), \pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{\Delta}\right)\right\rangle - \left\langle \pi_{E_{\boldsymbol{t}}}(\overline{w}), \pi_{E_{\boldsymbol{t}}}\left(\overline{\Delta}\right)\right\rangle\right| &\le 2\|\overline{\Delta}\|\|\overline{w}\|.
\end{aligned}$$

Note that $\overline{w} = z - \sum_{r=1}^{p} \overline{\phi}_{r,n} B^r z$ thus by the triangle inequality

$$\|\overline{w}\| \le \|z\| + \sum_{r=1}^{p} |\overline{\phi}_{r,n}| \|B^r z\| .$$

Since $|\overline{\phi}_{r,n}| = O_P(1)$ for all $1 \le r \le p$, we deduce from Lemma 6.1 that $\|\overline{w}\| = O_P\left(n^{1/2}\right)$. Since, by Lemma 6.9, $\|\overline{\Delta}\| = O_P(1)$, we obtain that

$$\sup_{\boldsymbol{t}}\left\{J_n\left(\overline{\Delta},\boldsymbol{t}\right) - \frac{2}{n}\left(\left\langle\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{\Delta}\right)\right\rangle - \left\langle\pi_{E_{\boldsymbol{t}}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}}}\left(\overline{\Delta}\right)\right\rangle\right)\right\} = O_P\left(n^{-1/2}\right). \tag{53}$$

For $0 < \alpha < \Delta_{\boldsymbol{\tau}^\star}$, using (31) and (41), we get:

$$\begin{aligned}
P\left(\left\|\bar{\boldsymbol{t}}_n - \boldsymbol{t}^\star\right\|_\infty \ge \alpha\right) &\le P\left(\min_{\boldsymbol{t}\in\mathcal{C}_{n,\alpha}} J_n\left(\overline{v},\boldsymbol{t}\right) \le 0\right) \\
&\le P\Bigg(\min_{\boldsymbol{t}\in\mathcal{C}_{n,\alpha}} \Big\{J_n\left(\overline{w},\boldsymbol{t}\right) + J_n\left(\overline{\Delta},\boldsymbol{t}\right) \\
&\quad - \frac{2}{n}\left(\left\langle\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{\Delta}\right)\right\rangle - \left\langle\pi_{E_{\boldsymbol{t}}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}}}\left(\overline{\Delta}\right)\right\rangle\right)\Big\} \le 0\Bigg) \\
&\le P\Bigg(\min_{\boldsymbol{t}\in\mathcal{C}_{n,\alpha}} \Big\{K_n\left(\overline{w},\boldsymbol{t}\right) + V_n\left(\overline{w},\boldsymbol{t}\right) + W_n\left(\overline{w},\boldsymbol{t}\right) + J_n\left(\overline{\Delta},\boldsymbol{t}\right) \\
&\quad - \frac{2}{n}\left(\left\langle\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{\Delta}\right)\right\rangle - \left\langle\pi_{E_{\boldsymbol{t}}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}}}\left(\overline{\Delta}\right)\right\rangle\right)\Big\} \le 0\Bigg) \\
&\le P\left(\min_{\boldsymbol{t}\in\mathcal{C}_{n,\alpha}} \left\{\frac{1}{2}K_n\left(\overline{w},\boldsymbol{t}\right) + V_n\left(\overline{w},\boldsymbol{t}\right) + W_n\left(\overline{w},\boldsymbol{t}\right)\right\} \le 0\right) \\
&\quad + P\Bigg(\min_{\boldsymbol{t}\in\mathcal{C}_{n,\alpha}} \Big\{\frac{1}{2}K_n\left(\overline{w},\boldsymbol{t}\right) + J_n\left(\overline{\Delta},\boldsymbol{t}\right) \\
&\quad - \frac{2}{n}\left(\left\langle\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{\Delta}\right)\right\rangle - \left\langle\pi_{E_{\boldsymbol{t}}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}}}\left(\overline{\Delta}\right)\right\rangle\right)\Big\} \le 0\Bigg).
\end{aligned}$$

Following the proof of Lemma 6.3, one can prove that

$$P\left(\min_{\boldsymbol{t}\in\mathcal{C}_{n,\alpha}} \left\{\frac{1}{2}K_n\left(\overline{w},\boldsymbol{t}\right) + V_n\left(\overline{w},\boldsymbol{t}\right) + W_n\left(\overline{w},\boldsymbol{t}\right)\right\} \le 0\right) \underset{n\to\infty}{\longrightarrow} 0\,.$$

Using (38), we get that

$$\begin{aligned}
&P\left(\min_{\boldsymbol{t}\in\mathcal{C}_{n,\alpha}} \left\{\frac{1}{2}K_n\left(\overline{w},\boldsymbol{t}\right) + J_n\left(\overline{\Delta},\boldsymbol{t}\right) - \frac{2}{n}\left(\left\langle\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{\Delta}\right)\right\rangle - \left\langle\pi_{E_{\boldsymbol{t}}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}}}\left(\overline{\Delta}\right)\right\rangle\right)\right\} \le 0\right) \\
&\quad \le P\left(\frac{1}{2}\underline{\lambda}^2\alpha + \min_{\boldsymbol{t}\in\mathcal{C}_{n,\alpha}} \left\{J_n\left(\overline{\Delta},\boldsymbol{t}\right) - \frac{2}{n}\left(\left\langle\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{\Delta}\right)\right\rangle - \left\langle\pi_{E_{\boldsymbol{t}}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}}}\left(\overline{\Delta}\right)\right\rangle\right)\right\} \le 0\right)
\end{aligned}$$

which goes to zero when n goes to infinity by (53).

Then Lemma 6.3 still holds if y is defined by (1). To show the rate of convergence, we use the same decomposition. As in the proof of Lemma 6.6, $P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})} J_n\left(\overline{v},\boldsymbol{t}\right) \le 0\right) \underset{n\to\infty}{\longrightarrow} 0$ for all $\alpha > 0$ and $0 < \gamma < 1/2$ is a sufficient condition for proving that $P\left(\widehat{\boldsymbol{t}}_n(y,\overline{\rho}_n) \in \mathcal{C}_{\alpha,\gamma,n}\right) \longrightarrow_{n\to\infty} 0$, which allows us to conclude on the rate of convergence of the estimated change-points. Note that

$$\begin{aligned}
P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}(\mathcal{I})} J_n\left(\overline{v},\boldsymbol{t}\right) \le 0\right) &\le P\left(\min_{\boldsymbol{t}\in\mathcal{C}'_{\alpha,\gamma,n}} \left\{\frac{1}{2}K_n\left(\overline{w},\boldsymbol{t}\right) + V_n\left(\overline{w},\boldsymbol{t}\right) + W_n\left(\overline{w},\boldsymbol{t}\right)\right\} \le 0\right) \\
&+ P\Bigg(\frac{1}{2}\underline{\lambda}^2\alpha + J_n\left(\overline{\Delta},\boldsymbol{t}\right) \\
&- \frac{2}{n}\left(\left\langle\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}_n^\star}}\left(\overline{\Delta}\right)\right\rangle - \left\langle\pi_{E_{\boldsymbol{t}}}\left(\overline{w}\right),\pi_{E_{\boldsymbol{t}}}\left(\overline{\Delta}\right)\right\rangle\right) \le 0\Bigg).
\end{aligned}$$

In the latter equation, the second term of the right-hand side goes to zero as n goes to infinity by (53). The first term of the right-hand side goes to zero when n goes to infinity by following the same line of reasoning as the one of Lemma 6.5. This concludes the proof of Proposition 3.2. $\square$

7 Tables and figures

n	7200					14400				
estimate \number of changes	<5	5	**6**	7	>7	<5	5	**6**	7	>7
$\widehat{m}_Y^0$	0	0	**100**	0	0	0	0	**100**	0	0
$\widehat{m}^0$	0	0	**100**	0	0	0	0	**100**	0	0
$\widehat{m}$	0	0	**97**	3	0	0	0	**100**	0	0
$\widehat{m}_{PP}$	0	0	**99**	1	0	0	0	**100**	0	0
$\widehat{m}^\star$	0	0	**97**	3	0	0	0	**100**	0	0
$\widehat{m}_{PP}^\star$	0	0	**99**	1	0	0	0	**100**	0	0
$\widehat{m}'$	0	0	**97**	3	0	0	0	**100**	0	0
$\widehat{m}_{PP}'$	0	0	**99**	1	0	0	0	**100**	0	0
estimate \order of the autoregression	0	1	**2**	3	>3	0	1	**2**	3	>3
$\widehat{p}'$	0	0	**96**	4	0	0	0	**97**	1	2
$\widetilde{\phi}_{n,1}^{(2)}$ RMSE	$1.99 \cdot 10^{-2}$					$1.64 \cdot 10^{-2}$				
$\widetilde{\phi}_{n,2}^{(2)}$ RMSE	$1.80 \cdot 10^{-2}$					$1.54 \cdot 10^{-2}$				

Table 1: Estimates of the number of changes, of the order of the autoregression, and RMSEs of the estimates of the autoregression parameters, for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (-1.2, -0.4, 0.4)$.

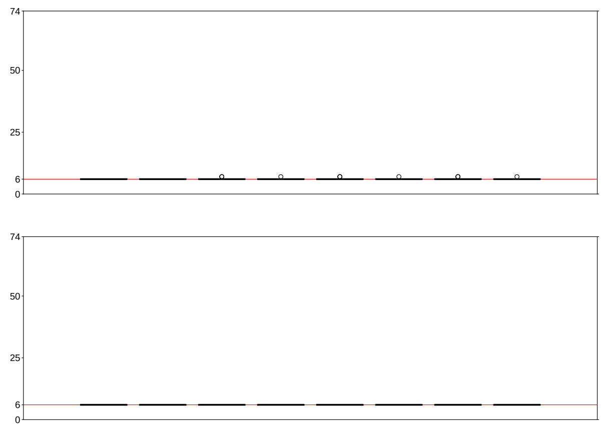

Figure 1: Boxplots of the estimates of the number of changes for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (-1.2, -0.4, 0.4)$. $n = 7200$ (top) or 14400 (bottom). In each plot, the estimates boxplots are in the following order (from left to right): $\widehat{m}_Y^0$, $\widehat{m}^0$, $\widehat{m}$, $\widehat{m}_{PP}$, $\widehat{m}^\star$, $\widehat{m}_{PP}^\star$, $\widehat{m}'$, $\widehat{m}_{PP}'$. The true number of changes is equal to 6 (red horizontal line).

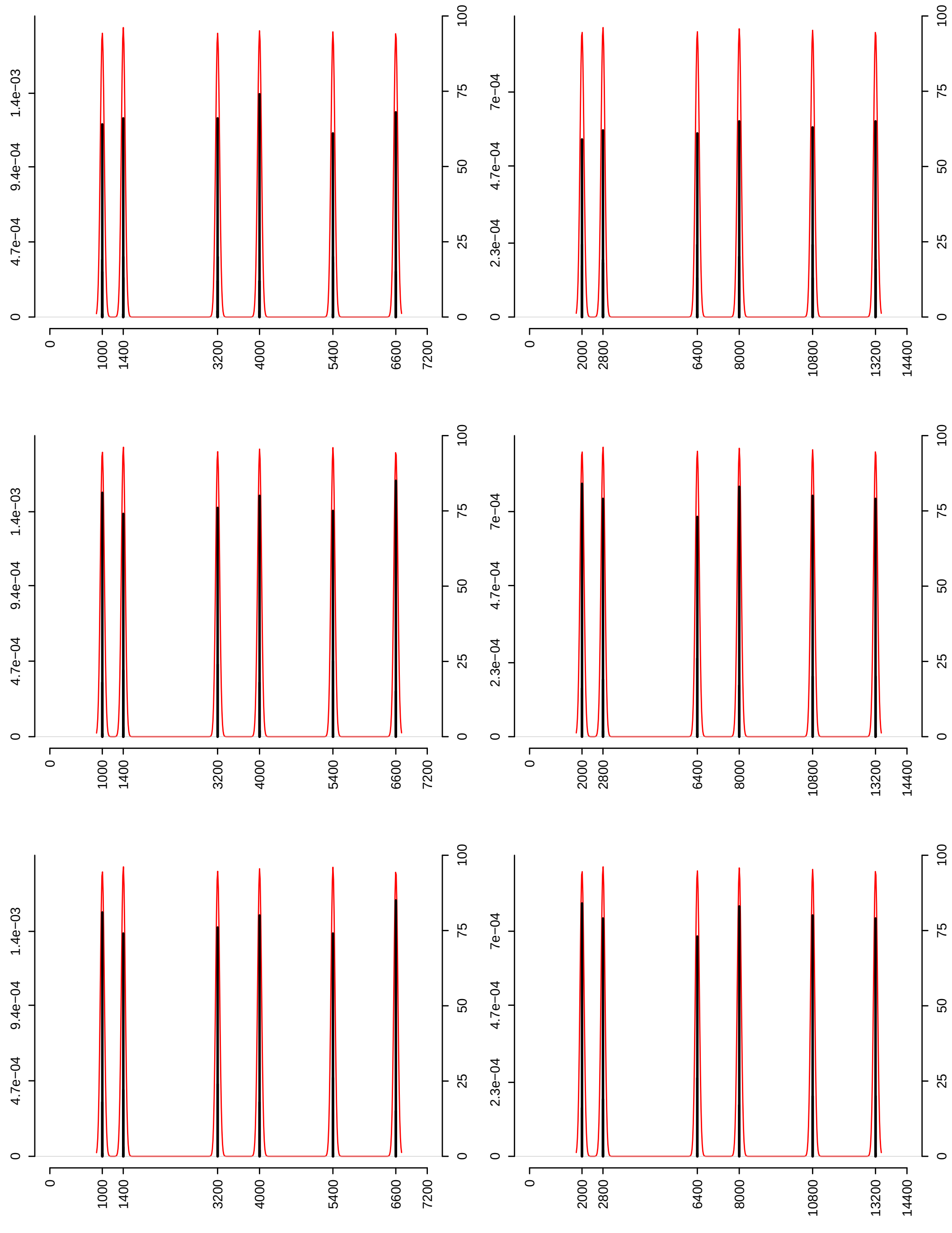

Figure 2: Frequency plots of the change-point location estimates for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (-1.2, -0.4, 0.4)$. $n = 7200$ (left) or 14400 (right). Estimates: $\widehat{t}_n^0$ (top), $\widehat{t}_{n,PP}$ (middle), $\widehat{t}'_{n,PP}$ (bottom). The black line represents the absolute frequency of each location between 1 and n in estimates (scale on right axis). The red line represents the Gaussian kernel density estimate of this dataset (scale on left axis).

n	7200					14400				
estimate \number of changes	<5	5	**6**	7	>7	<5	5	**6**	7	>7
$\widehat{m}_Y^0$	0	0	**0**	0	100	0	0	**0**	0	100
$\widehat{m}^0$	0	0	**0**	0	100	0	0	**0**	1	99
$\widehat{m}$	3	0	**91**	6	0	0	0	**99**	1	0
$\widehat{m}_{PP}$	3	0	**97**	0	0	0	0	**100**	0	0
$\widehat{m}^\star$	0	0	**94**	6	0	0	0	**99**	1	0
$\widehat{m}^\star_{PP}$	0	0	**100**	0	0	0	0	**100**	0	0
$\widehat{m}'$	4	0	**90**	6	0	0	0	**98**	2	0
$\widehat{m}'_{PP}$	4	0	**96**	0	0	0	0	**99**	1	0
estimate \order of the autoregression	0	1	**2**	3	>3	0	1	**2**	3	>3
$\widetilde{p}'$	0	0	**46**	24	30	0	0	**55**	15	30
$\widetilde{\phi}^{(2)}_{n,1}$ RMSE	$4.93 \cdot 10^{-2}$					$3.46 \cdot 10^{-2}$				
$\widetilde{\phi}^{(2)}_{n,2}$ RMSE	$3.13 \cdot 10^{-2}$					$2.16 \cdot 10^{-2}$				

Table 2: Estimates of the number of changes, of the order of the autoregression, and RMSEs of the estimates of the autoregression parameters, for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (1.6, -0.8, 0.4)$.

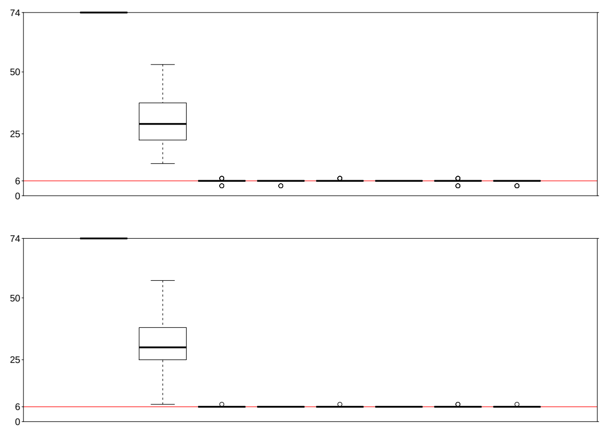

Figure 3: Boxplots of the estimates of the number of changes for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (1.6, -0.8, 0.4)$. $n = 7200$ (top) or 14400 (bottom). In each plot, the estimates boxplots are in the following order (from left to right): $\widehat{m}_Y^0$, $\widehat{m}^0$, $\widehat{m}$, $\widehat{m}_{PP}$, $\widehat{m}^\star$, $\widehat{m}^\star_{PP}$, $\widehat{m}'$, $\widehat{m}'_{PP}$. The true number of changes is equal to 6 (red horizontal line).

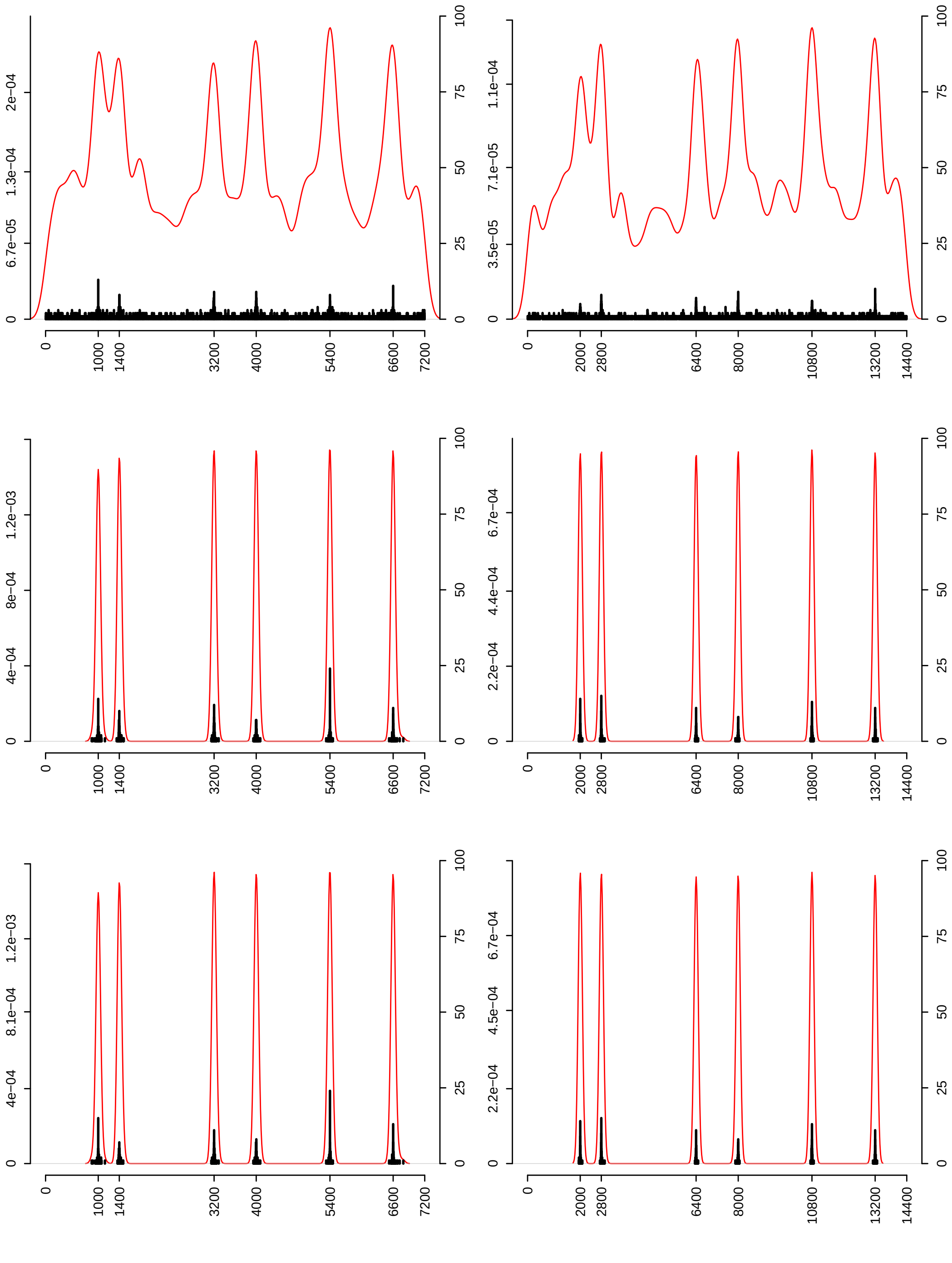

Figure 4: Frequency plots of the change-point location estimates for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (1.6, -0.8, 0.4)$. $n = 7200$ (left) or 14400 (right). Estimates: $\widehat{t}_n^0$ (top), $\widehat{t}_{n,PP}$ (middle), $\widehat{t}'_{n,PP}$ (bottom). The black line represents the absolute frequency of each location between 1 and n in estimates (scale on right axis). The red line represents the Gaussian kernel density estimate of this dataset (scale on left axis).

n	7200					14400				
estimate \number of changes	<5	5	**6**	7	>7	<5	5	**6**	7	>7
$\widehat{m}_Y^0$	0	0	**0**	0	100	0	0	**0**	0	100
$\widehat{m}^0$	0	0	**51**	20	29	0	0	**60**	17	23
$\widehat{m}$	3	0	**96**	3	1	0	0	**98**	2	0
$\widehat{m}_{PP}$	3	0	**97**	2	1	0	0	**98**	2	0
$\widehat{m}^\star$	0	0	**97**	3	0	0	0	**99**	1	0
$\widehat{m}^\star_{PP}$	0	0	**98**	2	0	0	0	**99**	1	0
$\widehat{m}'$	0	0	**97**	3	0	0	0	**99**	1	0
$\widehat{m}'_{PP}$	0	0	**98**	2	0	0	0	**99**	1	0
estimate \order of the autoregression	0	1	**2**	3	>3	0	1	**2**	3	>3
$\widehat{p}'$	0	0	**59**	25	16	0	0	**45**	30	25
$\widetilde{\phi}^{(2)}_{n,1}$ RMSE	$7.00 \cdot 10^{-2}$					$6.44 \cdot 10^{-2}$				
$\widetilde{\phi}^{(2)}_{n,2}$ RMSE	$4.20 \cdot 10^{-2}$					$3.68 \cdot 10^{-2}$				

Table 3: Estimates of the number of changes, of the order of the autoregression, and RMSEs of the estimates of the autoregression parameters, for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (0.2, 0.2, 0.4)$.

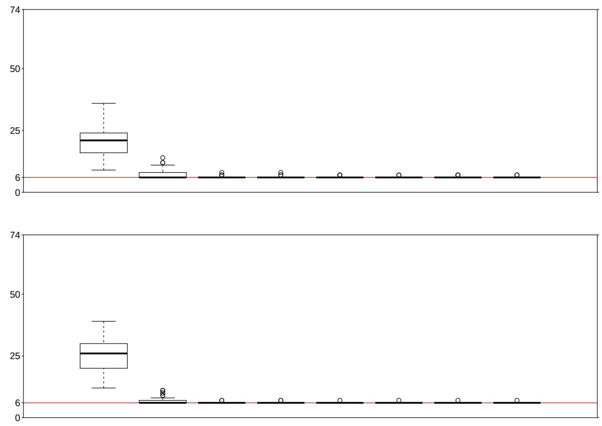

Figure 5: Boxplots of the estimates of the number of changes for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (0.2, 0.2, 0.4)$. $n = 7200$ (top) or 14400 (bottom). In each plot, the estimates boxplots are in the following order (from left to right): $\widehat{m}_Y^0$, $\widehat{m}^0$, $\widehat{m}$, $\widehat{m}_{PP}$, $\widehat{m}^\star$, $\widehat{m}^\star_{PP}$, $\widehat{m}'$, $\widehat{m}'_{PP}$. The true number of changes is equal to 6 (red horizontal line).

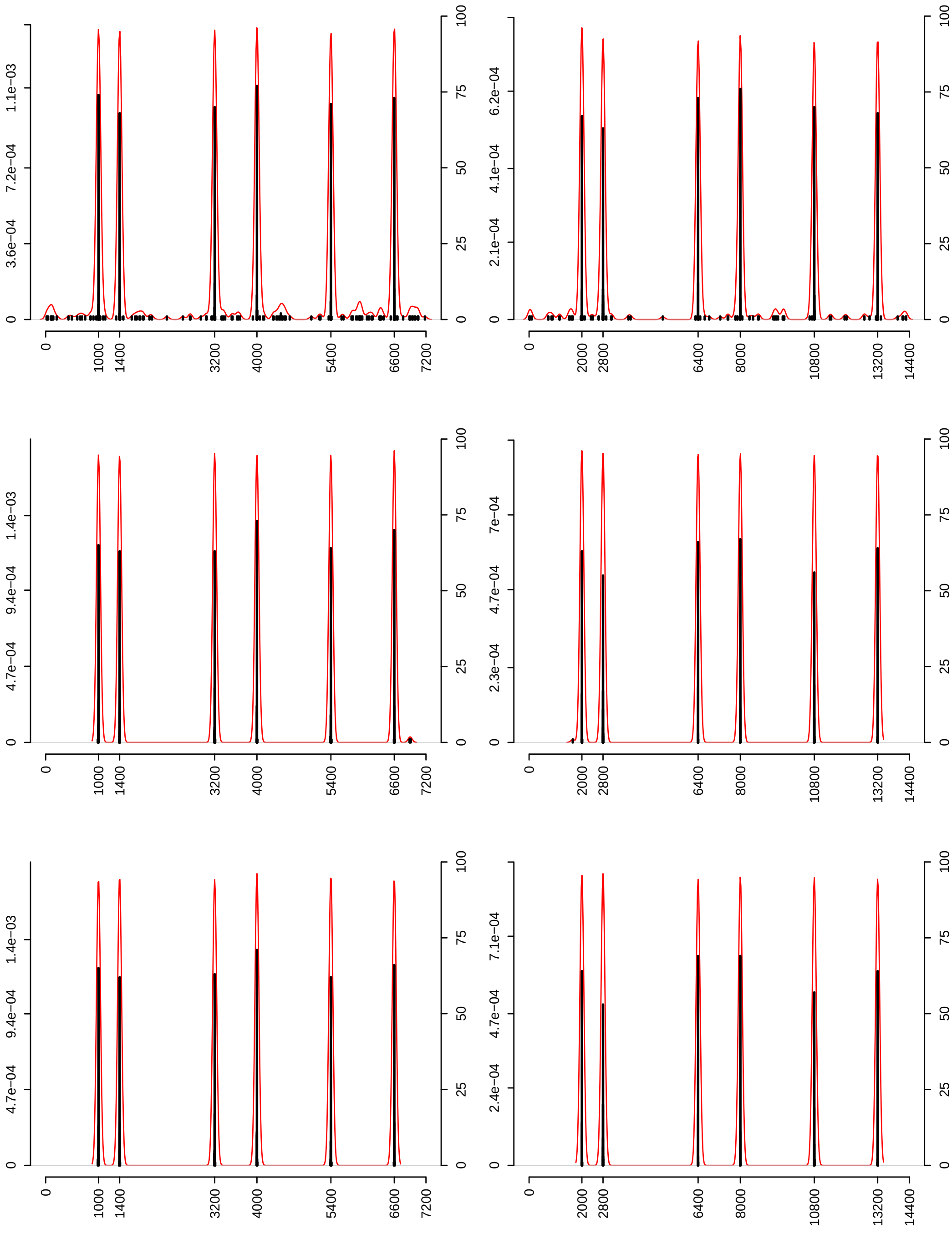

Figure 6: Frequency plots of the change-point location estimates for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (0.2, 0.2, 0.4)$. $n = 7200$ (left) or 14400 (right). Estimates: $\widehat{t}_n^0$ (top), $\widehat{t}_{n,PP}$ (middle), $\widehat{t}'_{n,PP}$ (bottom). The black line represents the absolute frequency of each location between 1 and n in estimates (scale on right axis). The red line represents the Gaussian kernel density estimate of this dataset (scale on left axis).

n	7200					14400				
estimate \number of changes	< 5	5	**6**	7	> 7	< 5	5	**6**	7	> 7
$\widehat{m}_Y^0$	0	0	**0**	0	100	0	0	**0**	0	100
$\widehat{m}^0$	0	0	**0**	0	100	0	0	**0**	0	100
$\widehat{m}$	13	0	**27**	4	56	25	0	**29**	9	37
$\widehat{m}_{PP}$	13	0	**28**	3	56	25	0	**33**	5	37
$\widehat{m}^\star$	0	1	**80**	18	1	0	0	**88**	11	1
$\widehat{m}_{PP}^\star$	1	0	**91**	7	1	0	0	**98**	2	0
$\widehat{m}'$	22	0	**54**	15	9	6	0	**80**	11	3
$\widehat{m}'_{PP}$	22	0	**66**	4	8	6	0	**90**	3	1
estimate \order of the autoregression	0	1	**2**	3	> 3	0	1	**2**	3	> 3
$\widehat{p}'$	0	0	**18**	12	70	0	0	**14**	12	74
$\widetilde{\phi}_{n,1}^{(2)}$ RMSE	$3.44 \cdot 10^{-1}$					$2.40 \cdot 10^{-1}$				
$\widetilde{\phi}_{n,2}^{(2)}$ RMSE	$2.41 \cdot 10^{-1}$					$1.71 \cdot 10^{-1}$				

Table 4: Estimates of the number of changes, of the order of the autoregression, and RMSEs of the estimates of the autoregression parameters, for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (0.2, 0.6, 0.4)$.

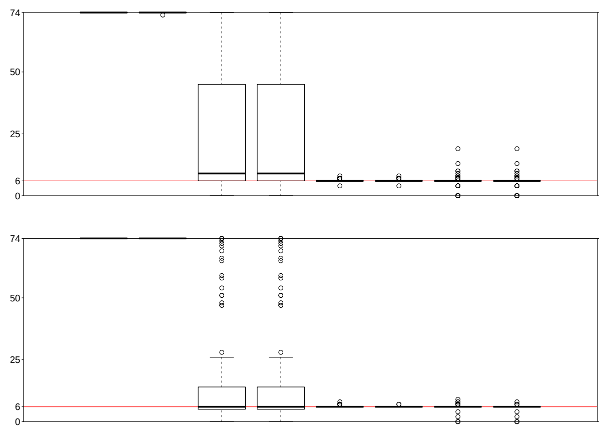

Figure 7: Boxplots of the estimates of the number of changes for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (0.2, 0.6, 0.4)$. $n = 7200$ (top) or 14400 (bottom). In each plot, the estimates boxplots are in the following order (from left to right): $\widehat{m}_Y^0$, $\widehat{m}^0$, $\widehat{m}$, $\widehat{m}_{PP}$, $\widehat{m}^\star$, $\widehat{m}_{PP}^\star$, $\widehat{m}'$, $\widehat{m}'_{PP}$. The true number of changes is equal to 6 (red horizontal line).

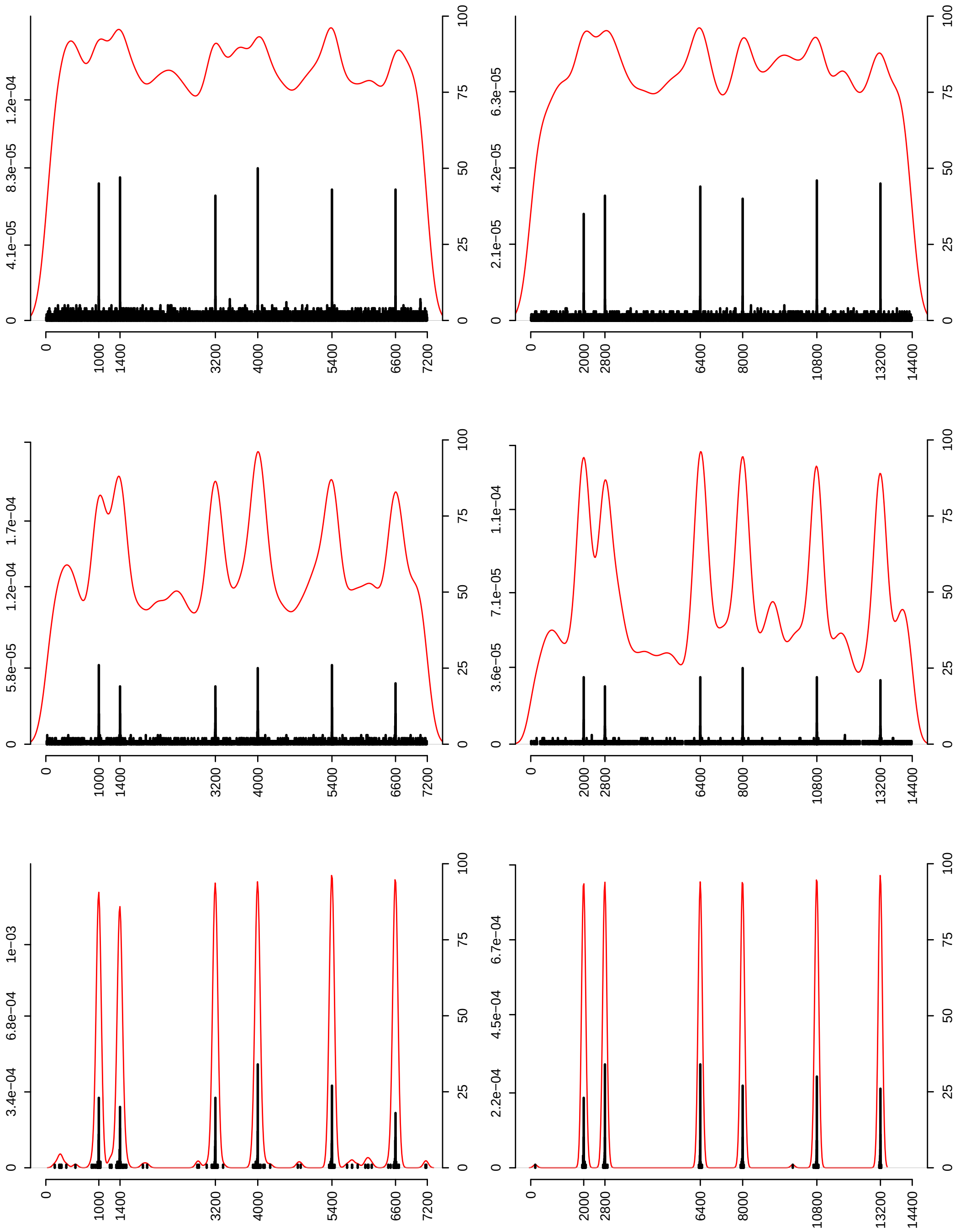

Figure 8: Frequency plots of the change-point location estimates for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (0.2, 0.6, 0.4)$. $n = 7200$ (left) or 14400 (right). Estimates: $\widehat{t}_n^0$ (top), $\widehat{t}_{n,PP}$ (middle), $\widehat{t}'_{n,PP}$ (bottom). The black line represents the absolute frequency of each location between 1 and n in estimates (scale on right axis). The red line represents the Gaussian kernel density estimate of this dataset (scale on left axis).

n	7200					14400				
estimate \number of changes	<5	5	**6**	7	>7	<5	5	**6**	7	>7
$\widehat{m}_Y^0$	0	0	**0**	0	100	0	0	**0**	0	100
$\widehat{m}^0$	0	0	**0**	0	100	0	0	**0**	0	100
$\widehat{m}$	2	0	**52**	28	18	1	0	**59**	28	12
$\widehat{m}_{PP}$	3	0	**85**	10	2	1	0	**96**	3	0
$\widehat{m}^\star$	0	0	**65**	28	7	0	0	**70**	29	1
$\widehat{m}_{PP}^\star$	0	0	**98**	2	0	0	0	**98**	2	0
$\widehat{m}'$	0	0	**63**	28	9	0	0	**74**	24	2
$\widehat{m}'_{PP}$	0	0	**100**	0	0	0	0	**99**	1	0
estimate \order of the autoregression	0	1	**2**	3	>3	0	1	**2**	3	>3
$\widehat{p}'$	0	0	**36**	20	44	0	0	**36**	21	43
$\widetilde{\phi}_{n,1}^{(2)}$ RMSE	$1.11 \cdot 10^{-1}$					$8.17 \cdot 10^{-2}$				
$\widetilde{\phi}_{n,2}^{(2)}$ RMSE	$5.16 \cdot 10^{-2}$					$3.76 \cdot 10^{-2}$				

Table 5: Estimates of the number of changes, of the order of the autoregression, and RMSEs of the estimates of the autoregression parameters, for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (0.4, 0.2, 0.2)$.

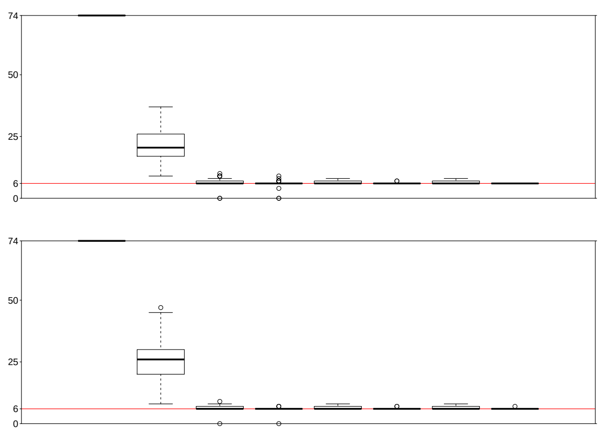

Figure 9: Boxplots of the estimates of the number of changes for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (0.4, 0.2, 0.2)$. $n = 7200$ (top) or 14400 (bottom). In each plot, the estimates boxplots are in the following order (from left to right): $\widehat{m}_Y^0$, $\widehat{m}^0$, $\widehat{m}$, $\widehat{m}_{PP}$, $\widehat{m}^\star$, $\widehat{m}_{PP}^\star$, $\widehat{m}'$, $\widehat{m}'_{PP}$. The true number of changes is equal to 6 (red horizontal line).

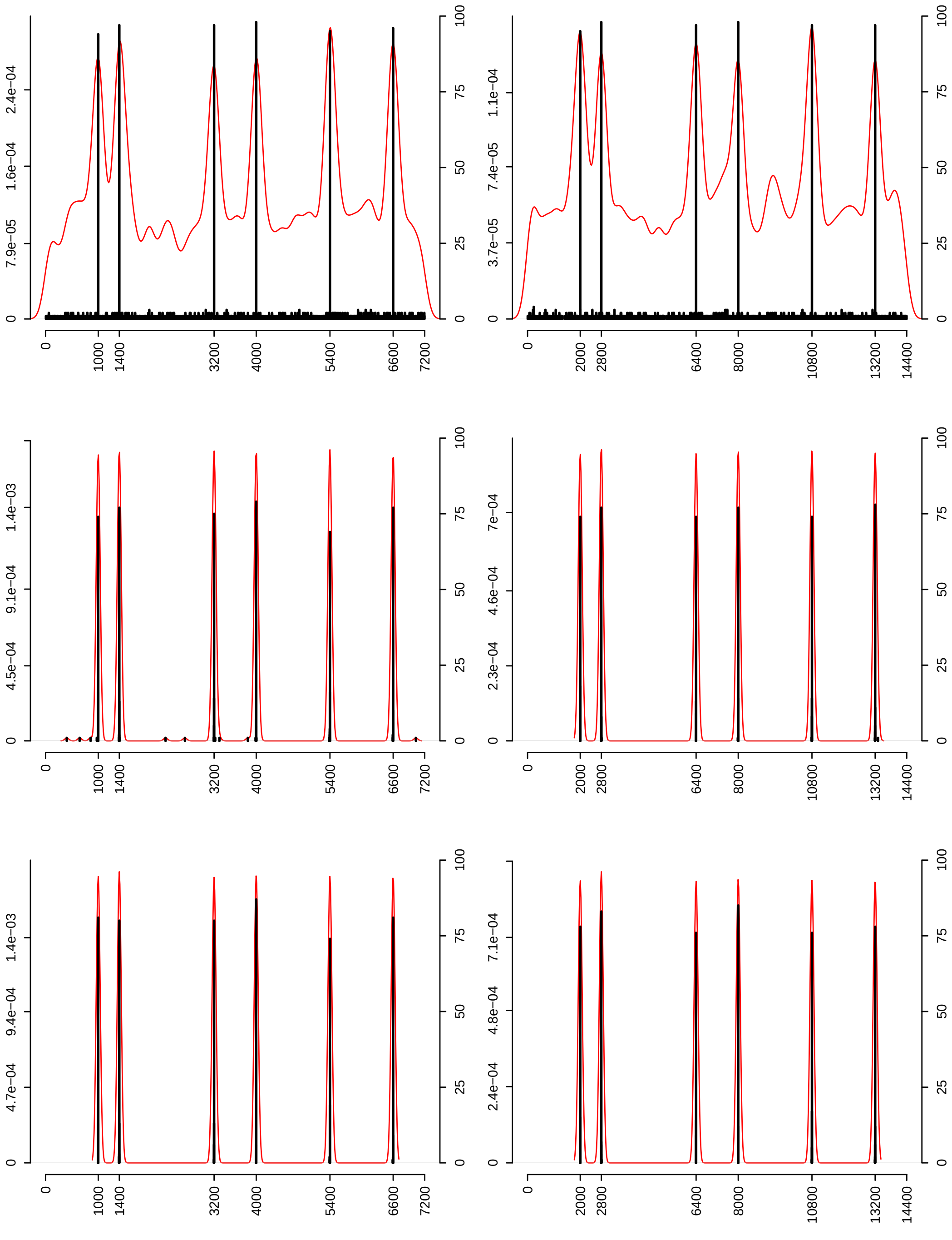

Figure 10: Frequency plots of the change-point location estimates for 100 AR(2) series with the parameters $(\phi_1^\star, \phi_2^\star, \sigma^\star) = (0.4, 0.2, 0.2)$. $n = 7200$ (left) or 14400 (right). Estimates: $\widehat{t}_n^0$ (top), $\widehat{t}_{n,PP}$ (middle), $\widehat{t'}_{n,PP}$ (bottom). The black line represents the absolute frequency of each location between 1 and n in estimates (scale on right axis). The red line represents the Gaussian kernel density estimate of this dataset (scale on left axis).

n	7200					14400				
estimate \number of changes	<5	5	**6**	7	>7	<5	5	**6**	7	>7
$\widehat{m}_Y^0$	0	0	**0**	0	100	0	0	**0**	0	100
$\widehat{m}^0$	0	0	**31**	16	53	0	0	**31**	14	55
$\widehat{m}$	2	0	**90**	4	4	0	0	**100**	0	0
$\widehat{m}_{PP}$	2	0	**92**	2	4	0	0	**100**	0	0
$\widehat{m}^\star$	0	0	**100**	0	0	0	0	**99**	1	0
$\widehat{m}_{PP}^\star$	0	0	**100**	0	0	0	0	**100**	0	0
$\widehat{m}'$	0	0	**99**	1	0	0	0	**99**	1	0
$\widehat{m}'_{PP}$	0	0	**99**	1	0	0	0	**100**	0	0
estimate \order of the autoregression	<4	4	**5**	6	>6	<4	4	**5**	6	>6
$\widetilde{p}'$	0	0	**41**	25	34	0	0	**45**	21	34
$\widetilde{\phi}_{n,1}^{(5)}$ RMSE	$1.01 \cdot 10^{-1}$					$6.92 \cdot 10^{-2}$				
$\widetilde{\phi}_{n,2}^{(5)}$ RMSE	$4.36 \cdot 10^{-2}$					$3.19 \cdot 10^{-2}$				
$\widetilde{\phi}_{n,3}^{(5)}$ RMSE	$3.54 \cdot 10^{-2}$					$2.45 \cdot 10^{-2}$				
$\widetilde{\phi}_{n,4}^{(5)}$ RMSE	$2.48 \cdot 10^{-2}$					$1.84 \cdot 10^{-2}$				
$\widetilde{\phi}_{n,5}^{(5)}$ RMSE	$3.72 \cdot 10^{-2}$					$2.35 \cdot 10^{-2}$				

Table 6: Estimates of the number of changes, of the order of the autoregression, and RMSEs of the estimates of the autoregression parameters, for 100 AR(5) series with the parameters $(\phi_1^\star, \phi_2^\star, \phi_3^\star, \phi_4^\star, \phi_5^\star, \sigma^\star) = (0.5, 0, 0, 0.5, -0.5, 0.4)$.

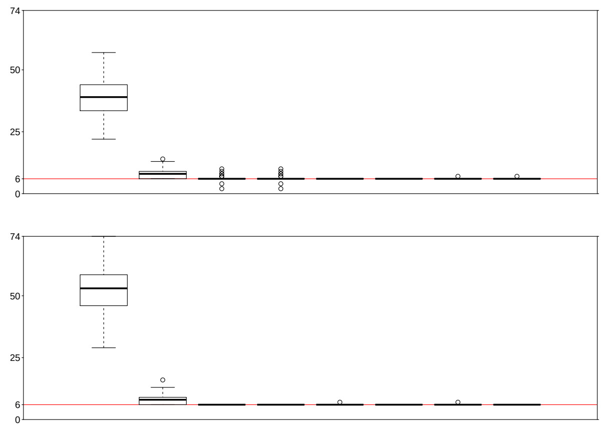

Figure 11: Boxplots of the estimates of the number of changes for 100 AR(5) series with the parameters $(\phi_1^\star, \phi_2^\star, \phi_3^\star, \phi_4^\star, \phi_5^\star, \sigma^\star) = (0.5, 0, 0, 0.5, -0.5, 0.4)$. $n = 7200$ (top) or 14400 (bottom). In each plot, the estimates boxplots are in the following order (from left to right): $\widehat{m}_Y^0$, $\widehat{m}^0$, $\widehat{m}$, $\widehat{m}_{PP}$, $\widehat{m}^\star$, $\widehat{m}_{PP}^\star$, $\widehat{m}'$, $\widehat{m}'_{PP}$. The true number of changes is equal to 6 (red horizontal line).

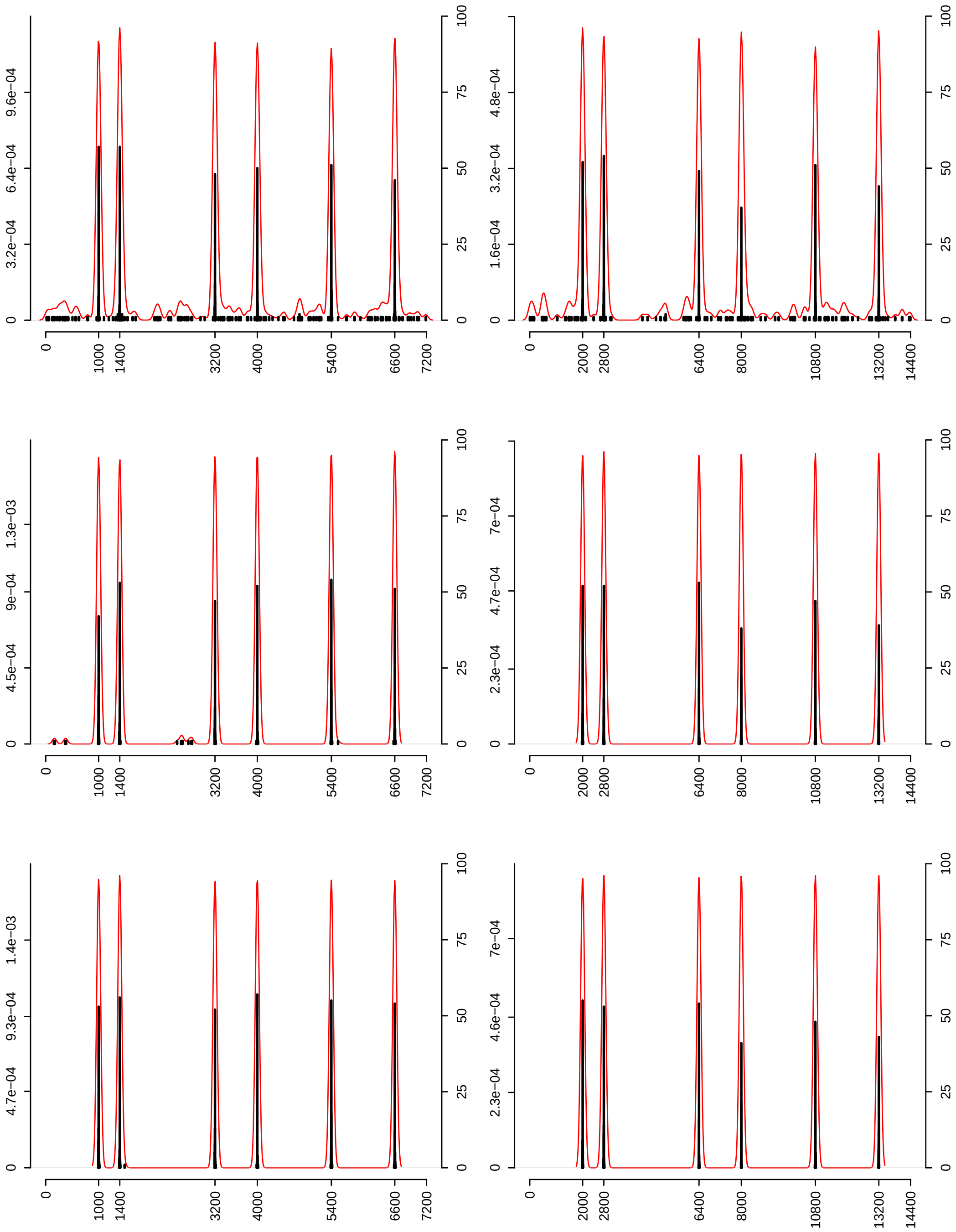

Figure 12: Frequency plots of the change-point location estimates for 100 AR(5) series with the parameters $(\phi_1^\star, \phi_2^\star, \phi_3^\star, \phi_4^\star, \phi_5^\star, \sigma^\star) = (0.5, 0, 0, 0.5, -0.5, 0.4)$. $n = 7200$ (left) or 14400 (right). Estimates: $\widehat{t}_n^0$ (top), $\widehat{t}_{n,PP}$ (middle), $\widehat{t}'_{n,PP}$ (bottom). The black line represents the absolute frequency of each location between 1 and n in estimates (scale on right axis). The red line represents the Gaussian kernel density estimate of this dataset (scale on left axis).

n	7200					14400				
estimate \number of changes	<5	5	**6**	7	>7	<5	5	**6**	7	>7
$\widehat{m}_Y^0$	0	0	**0**	0	100	0	0	**0**	0	100
$\widehat{m}^0$	0	0	**53**	5	42	0	0	**54**	2	44
$\widehat{m}$	0	0	**100**	0	0	0	0	**100**	0	0
$\widehat{m}_{PP}$	0	0	**100**	0	0	0	0	**100**	0	0
$\widehat{m}^\star$	0	0	**100**	0	0	0	0	**100**	0	0
$\widehat{m}^\star_{PP}$	0	0	**100**	0	0	0	0	**100**	0	0
$\widehat{m}'$	0	0	**100**	0	0	0	0	**100**	0	0
$\widehat{m}'_{PP}$	0	0	**100**	0	0	0	0	**100**	0	0
estimate \order of the autoregression	<4	4	**5**	6	>6	<4	4	**5**	6	>6
$\widehat{p}'$	0	0	**72**	17	11	0	0	**83**	12	5
$\widetilde{\phi}^{(5)}_{n,1}$ RMSE	$2.99 \cdot 10^{-2}$					$1.77 \cdot 10^{-2}$				
$\widetilde{\phi}^{(5)}_{n,2}$ RMSE	$1.24 \cdot 10^{-2}$					$1.05 \cdot 10^{-2}$				
$\widetilde{\phi}^{(5)}_{n,3}$ RMSE	$1.25 \cdot 10^{-2}$					$1.03 \cdot 10^{-2}$				
$\widetilde{\phi}^{(5)}_{n,4}$ RMSE	$1.28 \cdot 10^{-2}$					$1.01 \cdot 10^{-2}$				
$\widetilde{\phi}^{(5)}_{n,5}$ RMSE	$1.29 \cdot 10^{-2}$					$9.47 \cdot 10^{-3}$				

Table 7: Estimates of the number of changes, of the order of the autoregression, and RMSEs of the estimates of the autoregression parameters, for 100 AR(5) series with the parameters $(\phi_1^\star, \phi_2^\star, \phi_3^\star, \phi_4^\star, \phi_5^\star, \sigma^\star) = (0.5, 0, 0, 0, -0.5, 0.4)$.

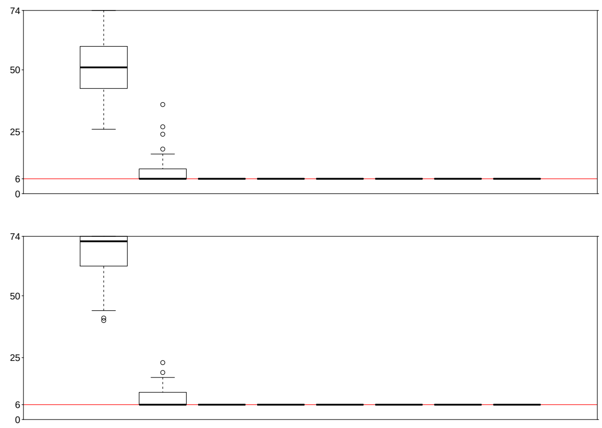

Figure 13: Boxplots of the estimates of the number of changes for 100 AR(5) series with the parameters $(\phi_1^\star, \phi_2^\star, \phi_3^\star, \phi_4^\star, \phi_5^\star, \sigma^\star) = (0.5, 0, 0, 0, -0.5, 0.4)$. $n = 7200$ (top) or 14400 (bottom). In each plot, the estimates boxplots are in the following order (from left to right): $\widehat{m}_Y^0$, $\widehat{m}^0$, $\widehat{m}$, $\widehat{m}_{PP}$, $\widehat{m}^\star$, $\widehat{m}^\star_{PP}$, $\widehat{m}'$, $\widehat{m}'_{PP}$. The true number of changes is equal to 6 (red horizontal line).

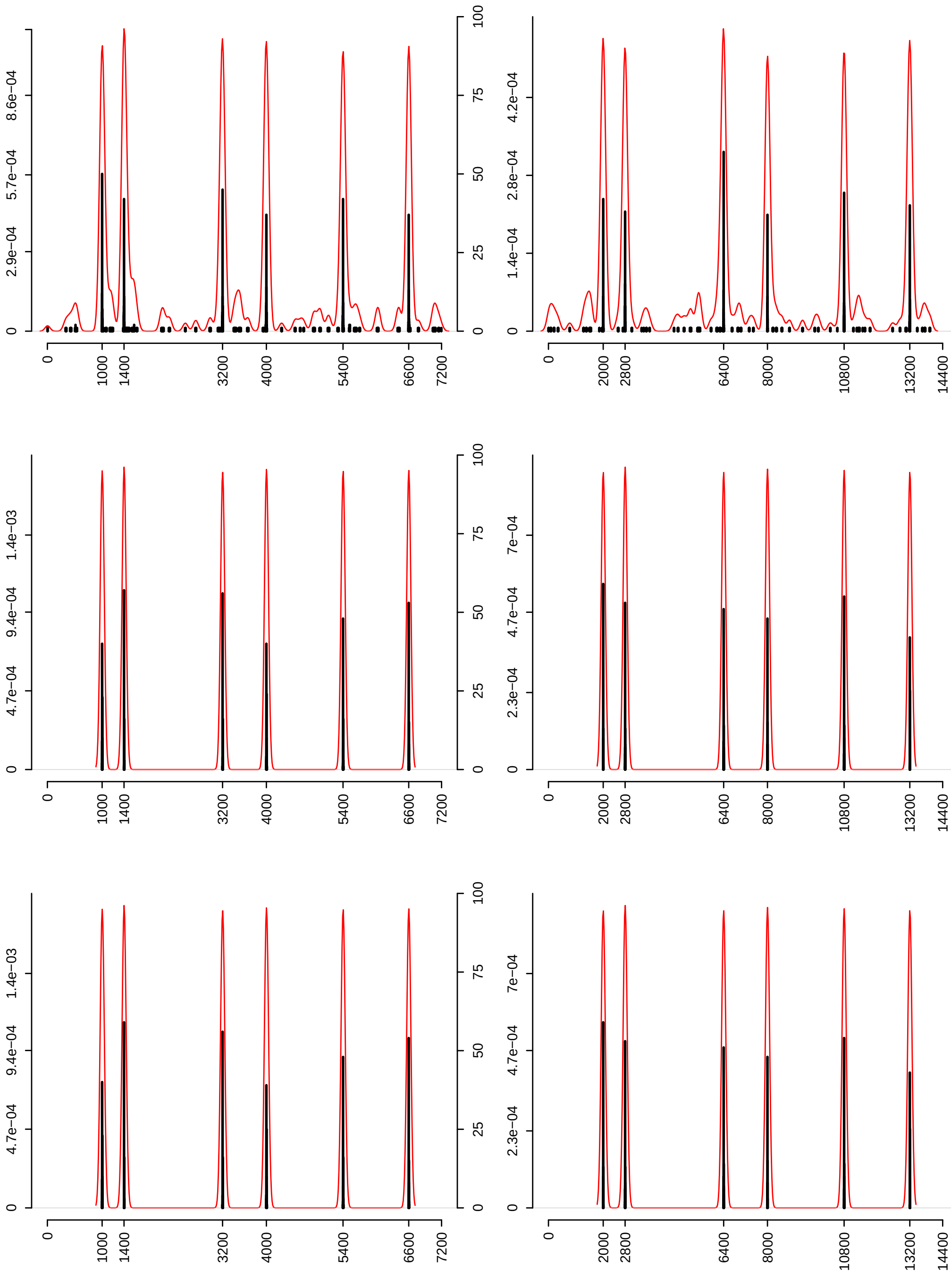

Figure 14: Frequency plots of the change-point location estimates for 100 AR(5) series with the parameters $(\phi_1^\star, \phi_2^\star, \phi_3^\star, \phi_4^\star, \phi_5^\star, \sigma^\star) = (0.5, 0, 0, 0, -0.5, 0.4)$. $n = 7200$ (left) or 14400 (right). Estimates: $\widehat{t}_n^0$ (top), $\widehat{t}_{n,PP}$ (middle), $\widehat{t}'_{n,PP}$ (bottom). The black line represents the absolute frequency of each location between 1 and n in estimates (scale on right axis). The red line represents the Gaussian kernel density estimate of this dataset (scale on left axis).

Acknowledgment

I thank Émilie Lebarbier, Céline Lévy-Leduc and Stéphane Robin for comments that greatly improved this paper.

References

Arcones, M. A. (1994). "Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors". In: *The Annals of Probability*, pp. 2242–2274.

Auger, I. E. and C. E. Lawrence (1989). "Algorithms for the optimal identification of segment neighborhoods". In: *Bulletin of Mathematical Biology* 51.1, pp. 39–54.

Azencott, R. and D. Dacunha-Castelle (1986). *Series of Irregular Observations. Forecasting and model building.* Red. by J. Gani and C. C. Heyde. Vol. 2. Applied Probability. New York: Springer New York.

Bardet, J.-M., W. C. Kengne, and O. Wintenberger (2012). "Detecting multiple change-points in general causal time series using penalized quasi-likelihood". In: *Electronic Journal of Statistics*, pp. 1–50.

Billingsley, P. (1995). *Probability and Measure.* Wiley Series in Probability and Statistics. Wiley.

Braun, J. V. and H.-G. Müller (1998). "Statistical methods for DNA sequence segmentation". In: *Statistical Science*, pp. 142–162.

Braun, J. V., R. K. Braun, and H.-G. Müller (2000). "Multiple changepoint fitting via quasi-likelihood, with application to DNA sequence segmentation". In: *Biometrika* 87.2, pp. 301–314.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods.* Springer series in statistics. Springer-Verlag.

Chakar, S., E. Lebarbier, C. Lévy-Leduc, and S. Robin (2014). "A robust approach for estimating change-points in the mean of an AR(1) process". In: *arXiv eprint 1403.1958.*

Frick, K., A. Munk, and H. Sieling (2014). "Multiscale change point inference". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76.3, pp. 495–580.

Gazeaux, J., S. Williams, M. King, M. Bos, R. Dach, M. Deo, A. W. Moore, L. Ostini, E. Petrie, M. Roggero, F. N. Teferle, G. Olivares, and F. H. Webb (2013). "Detecting offsets in GPS time series: First results from the detection of offsets in GPS experiment". In: *Journal of Geophysical Research (Solid Earth)* 118.5.

Killick, R., P. Fearnhead, and I. A. Eckley (2012). "Optimal detection of changepoints with a linear computational cost". In: *Journal of The American Statistical Association* 107.500, pp. 1590–1598.

Lavielle, M. (1999). "Detection of multiple changes in a sequence of dependent variables". In: *Stochastic Processes and their Applications* 83.1, pp. 79–102.

— (2005). "Using penalized contrasts for the change-point problem". In: *Signal Processing* 85.8, pp. 1501–1510.

Lavielle, M. and E. Moulines (2000). "Least-squares Estimation of an Unknown Number of Shifts in a Time Series". In: *Journal of Time Series Analysis* 21.1, pp. 33–59.

Lu, Q., R. Lund, and T. C. M. Lee (2010). "An MDL approach to the climate segmentation problem". In: *The Annals of Applied Statistics* 4.1, pp. 299–319.

Lévy-Leduc, C., H. Boistard, E. Moulines, M. S. Taqqu, and V. A. Reisen (2011). "Robust estimation of the scale and of the autocovariance function of Gaussian short-and long-range dependent processes". In: *Journal of Time Series Analysis* 32.2, pp. 135–156.

Ma, Y. and M. G. Genton (2000). "Highly Robust Estimation of the Autocovariance Function". In: *Journal of Time Series Analysis* 21.6, pp. 663–684.

Mestre, O. (2000). "Méthodes statistiques pour l'homogénéisation de longues séries climatiques". PhD thesis. Université Toulouse 3.

Picard, F., S. Robin, M. Lavielle, C. Vaisse, and J.-J. Daudin (2005). "A statistical approach for array CGH data analysis". In: *BMC bioinformatics* 6.1, p. 27.

Rigaill, G. (2010). "Pruned dynamic programming for optimal multiple change-point detection". In: *arXiv eprint 1004.0887*.

Rousseeuw, P. J. and C. Croux (1993). "Alternatives to the Median Absolute Deviation". In: *Journal of The American Statistical Association* 88.424, pp. 1273–1283.

Schwarz, G. (1978). "Estimating the dimension of a model". In: *The Annals of Statistics*, pp. 461–464.

Sheather, S. J. and M. C. Jones (1991). "A reliable data-based bandwidth selection method for kernel density estimation". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pp. 683–690.

Williams, S. (2003). "Offsets in Global Positioning System time series". In: *Journal of Geophysical Research (Solid Earth)* 108.

Yao, Y.-C. (1988). "Estimating the number of change-points via Schwarz' criterion". In: *Statistics & Probability Letters* 6.3, pp. 181–189.

Zhang, N. R. (2005). "Change-point detection and sequence alignment: statistical problems of genomics". PhD thesis. Stanford University.

Zhang, N. R. and D. O. Siegmund (2007). "A modified Bayes Information Criterion with applications to the analysis of Comparative Genomic Hybridization data". In: *Biometrics* 63.1, pp. 22–32.